ESTIMATING ENGINEERING AND MANUFACTURING DEVELOPMENT

COST RISK USING LOGISTIC AND MULTIPLE REGRESSION

THESIS

John V. Bielecki, Captain, USAF

AFIT/GCA/ENC/03-01

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/GCA/ENC/03-01

ESTIMATING ENGINEERING AND MANUFACTURING DEVELOPMENT COST RISK USING LOGISTIC AND MULTIPLE REGRESSION

THESIS

Presented to the Faculty

Department of Mathematics and Statistics

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Cost Analysis

John V. Bielecki, BS

Captain, USAF

March 2003

AFIT/GCA/ENC/03-01

ESTIMATING ENGINEERING AND MANUFACTURING DEVELOPMENT COST RISK USING LOGISTIC AND MULTIPLE REGRESSION

John V. Bielecki, BS
Captain, USAF

Approved:

//Signed//
Edward D. White, Major, USAF (Chairman) date

//Signed//
Michael A. Greiner, Major, USAF (Member) date

//Signed//
Michael J. Seibel (Member) date

Acknowledgements

I thank God, the country and the Pepsi Company® for helping me make it through this ordeal.

I would like to especially thank Maj Edward White, my thesis advisor, for all his patience with my repetitive questions and his guidance on this project. I extend a sincere thanks to my committee members, Mr. Michael Seibel and Maj Michael Greiner for their constructive guidance on this thesis. I also want to thank my friend Captain Vince Sipple for his invaluable assistance in helping complete this thesis and for the opportunity to follow-up on his original idea.

I would like to thank my wife for her encouragement, inspiration and infinite patience of this undertaking. Finally, thanks to my children, whose boundless energy kept me on track with what is truly important.

John V. Bielecki

Table of Contents

Page

List of Figures

List of Tables

AFIT/GCA/ENC/03-01

Abstract

Cost Growth in Department of Defense (DoD) major weapon systems has been an on-going problem for more than 30 years. Previous research has demonstrated the use of a two-step logistic and multiple regression methodology to predicting cost growth produces desirable results versus traditional single-step regression. This research effort validates, and further explores the use of a two-step procedure for assessing DoD major weapon system cost growth using historical data.

We compile programmatic data from the Selected Acquisition Reports (SARs) between 1990 and 2001 for programs covering all defense departments. Our analysis concentrates on cost growth in the research and development dollar accounts for the Engineering and Manufacturing Development phase of acquisition. We investigate the use of logistic regression in cost growth analysis to predict whether or not cost growth will occur in a program. If applicable, the multiple regression step is implemented to predict how much cost growth will occur. Our study focuses on four of the seven SAR cost growth categories within the research and development accounts – schedule, estimating, support, and other. We study each of these four categories individually for significant cost growth characteristics and develop predictive models for each.

ESTIMATING ENGINEERING AND MANUFACTURING DEVELOPMENT COST RISK USING LOGISTIC AND MULTIPLE REGRESSION

I. Introduction

General Issue

The Department of Defense (DoD) budget has been under intense Congressional scrutiny and downward pressure since the 1980's military build up under President Reagan. Part of this scrutiny is justified, as the price of most new major weapons system programs skyrocket past their original estimated cost. This increase in price, or cost growth, of major weapons system has averaged 20 percent over the past 30 years, according to a 1993 RAND study (Drezner, 1993:xiii-xiv). Inevitably, these unexpected increases in program costs have manifested in requests for supplemental funding from Congress for the respective program.

Today, the American public and Congress can no longer tolerate the persistent cost overruns on new weapons systems. The DoD budget has been reduced 29 percent over the last 16 years, and Congress has enacted legislation to monitor and control weapons system cost overruns. The legislation, passed in the 1980's, called the Nunn-McCurdy Act requires Congress to be notified of any program whose unit cost increases by 15 percent or more. Any unit cost increase of 25 percent or more requires Pentagon certification that the program is vital to national security to continue operations (Weinberger, 2002). Therefore, it is essential that DoD cost estimators (and program

managers) work to contain, and even reduce, the amount of cost growth exhibited by a weapon system.

Cost growth in the procurement of major weapon systems can be the attributed to poor program management or contractor inefficiencies, however, it mainly stems from risk and uncertainties about the program. The cost estimate must take into account not only the actual costs of the program under development but also the risks and uncertainties associated with the program. Cost growth is defined as the ratio of a weapon system's current estimate to some prior estimate, generally the Development Estimate (DE) (Hough, 1992:v). To control cost growth, managers must focus on accurately assigning dollar values to risks, so that the original estimate from which cost growth is calculated is more accurate (Sipple, 2002:2).

Specific Issue

Cost estimators use a wide range of methodologies when assigning values to risk elements in a weapons system cost estimate. The estimating methodology used is a function of the type of item being estimated and where the item is in the acquisition life cycle. Early in the life cycle, when uncertainty is greatest, the estimator will utilize an expert opinion or analogy methodology to establish a value on each element of a program. Individual elements are then summed to achieve the overall program estimate or baseline estimate. Analogy is simply valuing the new estimate on a similar existing or analogous system. These methods, as expected, are subjective and could be improved upon.

Later in the life cycle, the estimator will utilize historical or actual costs to value the program elements. This method is potentially more accurate because more information about the program is known and uncertainty is reduced. In this scenario, the baseline estimate is likely undervalued in terms of risk. An alternative, less subjective, method of valuing and forecasting program estimates must be used earlier in the life cycle to reduce the measured DoD cost growth.

Statistical regression methods have previously been proven effective in determining cost growth relationships, as well as, the ability to predict the amount of cost growth (Sipple, 2002:3). This research seeks to build upon the work of Sipple (2002) in providing cost estimators a model to effectively estimate risk earlier in the acquisition life cycle so that overall DoD cost growth can be reduced.

Scope and Limitations of the Study

The "Selected Acquisition Reports (SARs) are the primary means by which DoD reports the status of major acquisitions to Congress" (Jarvaise, 1996:3). They represent a vast collection of programmatic reports and data from which the majority of cost growth calculations are based (Hough, 1992:v). The SARs are widely available and contain relatively reliable data on cost growth. For these reasons, the SARs are the source of choice for cost growth analysis and the basis for our research. The SARs provides two estimates for each program: the baseline estimate (usually the DE) and the current estimate (most recent available). Additionally, the SARs breakdown each program's cost variance into seven categories: Economic, Quantity, Estimating, Engineering, Schedule, Support, and Other (Hough, 1992:5; Drezner, 1993:7). Any deviation from a program's

baseline is then calculated in terms of one of these seven cost variance categories and reported in base-year and then-year dollars to account for inflation. Comparisons can then be easily made between programs or over time.

Overall, the SARs contain nineteen sections with pertinent program data in each. These sections provide additional details that are essential to conducting cost growth analysis. In part, these sections include; mission and description, schedule and technical data, acquisition cost and variances, contract and production information, and a funding summary.

In this research, we measure cost growth as a percentage increase from the DE as listed in the SAR. This research will only focus on cost growth in the Research and Development, Test and Evaluation (RDT&E) accounts during the Engineering and Manufacturing Development (EMD) phase of acquisition. Since we are building upon previous research in this area, we omit study of the Engineering changes cost variance category since it has already been analyzed in Sipple's (2002) thesis. Additionally, we will not consider the categories of Economic and Quantity cost variances as these categories, by convention, are usually beyond the control of the cost estimator. Moreover, the usefulness of these areas to our research sponsor is negligible. Thus, we seek insight into what causes cost growth and the amount of cost growth we can expect from the remaining SAR categories: Estimating, Schedule, Support and Other.

Since this is a follow-on research study we continue with the originally defined guidelines established in Sipple's (2002) thesis. That is, this study is based on a database comprised of only programs that use the DE as the baseline estimate and programs whose EMD phase of acquisition falls within the period 1990-2001. Further, "only one SAR per

program is used, the most recent available, and in some instances, the most recent available DE-based SAR is the last SAR of the EMD phase" (Sipple, 2002:4).

The SARs do have limitations, but none that impede its use as our source of data in this research. However, there are some SAR limitations that further limit the scope of this research effort (e.g., security classification), and that some DEs may already contain some undisclosed monetary estimate of risk. This research will only use data from unclassified programs. Chapter III further describes these limitations.

This research is an extension of the innovative methodology used by Sipple (2002). Sipple's unique two-step methodology first utilizes logistic regression to predict which programs will have cost growth, and then second, uses multiple regression to predict the amount of cost growth that will occur. To the best of our knowledge, Sipple's (2002) research is the first (and only) documented use of logistic regression for predicting cost growth. Although, the use of multiple regression has been previously utilized to predict cost growth, the combination of the two together is on the forefront of the field.

Research Objectives

This study has three main objectives. First, use logistic regression to determine if certain program characteristics predict whether a program experiences cost growth in the RDT&E budget during the EMD phase of development. Logistic regression differs from multiple regression in that it predicts a binary response. In our case the binary response is: *Does a program experience cost growth, Yes or No?* Second, the study seeks to find predictors of which cost growth occurs. We use multiple regression to determine the amount (value) of cost growth in the RDT&E budget in the EMD phase of development.

Finally, we seek to develop a predictive model that may be used by cost estimators early in a programs acquisition life cycle to ascertain potential cost growth in the RDT&E budget in the EMD phase of program development (Sipple, 2002:5).

Chapter Summary

This research seeks to expand upon the cost estimating methodology developed in Sipple's (2002) thesis. The goal of this study is to provide cost estimators a model to effectively estimate and value risk earlier in a program's acquisition life cycle. The intent being, a reduction in the overall DoD cost growth rate from current levels. The methodology we use is a two-step process one, perform logistic regression on historical SARs to identify potential cost growth within a program and then two, use multiple regression to predict the amount of cost growth.

II. Literature Review

Chapter Overview

This chapter provides an overview of previous cost growth research. We begin with a synopsis of the DoD acquisition process and current operating environment. We continue with an analysis of risk and uncertainty factors that effect cost growth, and follow with a comprehensive review and discussion of pertinent cost growth research as it relates to ours. The knowledge and insight garnered from this literature review assists us in building a model that predicts RDT&E cost growth during the EMD phase of acquisition with the intent of reducing overall DoD cost growth.

The Acquisition Process

An awareness of the acquisition process is an important first step in understanding where cost growth occurs, and how it is measured. The Department of Defense Instruction (DoDI) 5000.2, Operation of the Defense Acquisition System, establishes the management framework, policy, and guidance for translating "mission needs" into major weapon systems acquisition programs. The process, officially known as the DoD Acquisition Process, consists of four milestones (otherwise known as decision points), four phases, and three activities (DoDI 5000.2). The four milestones are best recognized as: MS 0, MS I, MS II and MS III, however, a January 2001 change to DoDI 5000.2 reconfigures the four milestones to three milestones and renames them: A, B, and C. Since this research is based on data from the Selected Acquisition Reports (SAR) for programs having an EMD phase of development from 1990 – 2001, the old format and

terminology is used throughout this report. The four (old) phases of the acquisition process are: Phase 0 - Concept Exploration; Phase I – Program Definition and Risk Reduction, Phase II – Engineering and Manufacturing Development, and Phase III – Production, Fielding/Deployment, and Operational Support (DoDI 5000.2). The three activities are: Pre-System Acquisition, System Acquisition, and Sustainment.

A brief explanation of each of the milestones and phases is listed for clarity. The descriptions are taken from Howard Jaynes' 1999 thesis on *Correlation Analysis: Army Acquisition Program Cycle Time and Cost Variation*, which serves as an excellent source of clear, concise acquisition process information (Jaynes, 1999:11-13). See Jaynes for further details on the acquisition process.

- Milestone 0: conduct concept studies. Validation of the mission need and identification of possible alternatives. Approval of MS 0 by the Defense Acquisition Board (DAB) authorizes entry into Phase 0.
- Phase 0: Concept Exploration (CE). The mission need and the alternatives are further defined in terms of cost, schedule, and performance objects. Costs are incorporated in the Acquisition Program Baseline (APB). Acquisition Strategies are developed and the Operation Requirements Documents (ORD) is prepared.
- Milestone I: official approval to begin a new program.
- Phase I: Program Definition and Risk Reduction (PDRR). The program is defined in terms of designs and technological approaches. Prototyping and early operational assessments are used to reduce risk. Identification of cost and schedule trade-offs.

- Milestone II: approval to enter Phase II. The Milestone Decision Authority (MDA) evaluates the acquisition strategy and updated APB (development baseline) of the program before authorizing continuation. Note: this is the estimate we use in our research to calculate cost growth.
- Phase II: Engineering and Manufacturing Development. The program is transformed into a cost-effective, stable design. Developmental testing is conducted to ensure performance capabilities are satisfied and Low Rate Initial Production (LRIP) is authorized to further validate the new system.
- Milestone III: approval to enter Phase III. MDA reviews the acquisition strategy and updated APB (production baseline) program before approving entry in Phase III.
- Phase III: Production, Fielding/Deployment and Operational Support. The program enters full rate production and works to achieve Initial Operational Capability (IOC). IOC is the first deployment of a weapons system to an operational unit.

The first step in building a model to predict cost growth is to define a method for computing cost growth. Within the DoD, there are several methodologies for calculating cost growth, with the main difference being "the purpose or objective" of the analysis being conducted (Calcutt, 1993, 7-8). Cost growth generally refers to the difference (in price) between a program's inception or initial estimate and the most recent or final total estimate of cost for an acquisition program (Hough, 1992:10).

Our research continues with the originally defined cost growth computation established by Sipple (2002). Which defines cost growth as the percentage price increase

from the Development Estimate (DE) to the most recent available current estimate as listed in the SAR (Sipple, 2002:3). Figure 1 depicts where the DE fits into the acquisition framework.

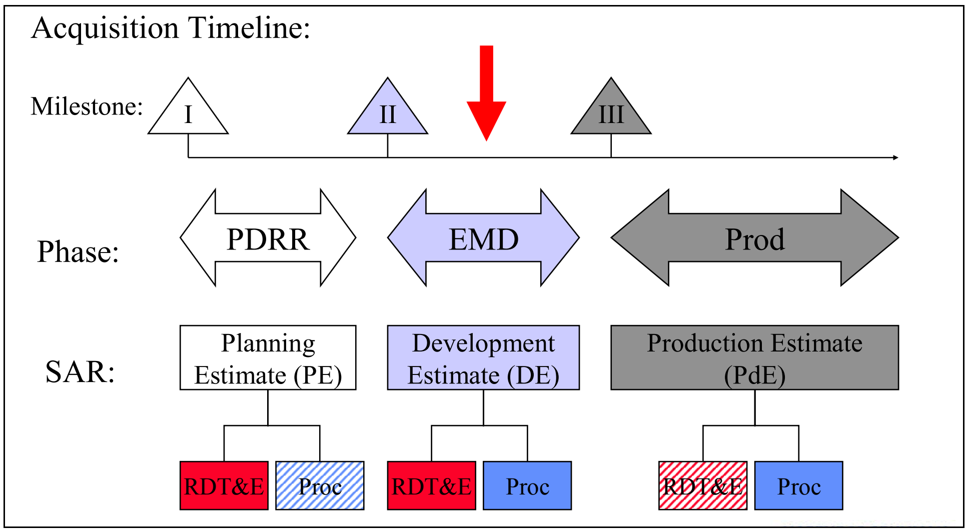


Figure 1 - Acquisition Timeline (Dameron, 2001:4)

The Environment

We now explore some of the environmental factors that influence cost growth. Since the fall of the Berlin Wall, the DoD budget has been under ever increasing downward pressure, falling from a high of $418.4 billion in 1985 to $296.3 in 2001 billion (29.18%) (Jaynes, 1999:4). All levels of the DoD structure feel the effects of this decline. Doing more with less is the daily mantra, particularly within a major weapons system program office. Moreover, weapons programs with exorbitant cost growth during this period of reduced funding, have garnered harsh Congressional and Presidential attention. For example, in January 1991(then) Secretary of Defense Cheney cancelled the

Navy A-12 program after costs inexplicably skyrocketed and "no one could tell him the program's final cost" (Christensen, 2002:105).

In January 2002, President Bush renewed emphasis on "realistic costing" as a way to control spiraling defense spending in this austere funding environment (Grossman, 2002:1). The idea of "realistic costing" is not new. Administrations in the early eighties also advocated realistic costing as a means to control spending (Sipple, 2002:9). Realistic costing recognizes that many programs routinely underestimate the true cost of development; they low-ball their initial price to get funding and then once funded, lobby for upward adjustments to cover true costs (Weinberger, 2002). Realistic costing implies that if program estimate were more realistic, hence more accurate, cost growth would be contained. Given such an austere funding environment, and the current political scrutiny, we conclude there is considerable pressure to deliver more accurate cost estimates within DoD. Our research seeks to satisfy this need for realistic cost estimates.

Risk and Uncertainty

What exactly do 'risk' and 'uncertainty' mean, and how do they relate to cost growth? According to a PricewaterhouseCoopers guide on *Uncertainty and Risk*, "the word 'uncertainty' means a number of different values can exist (Rodgers, 199:1). 'Risk' means the possibility of loss or gain as a result of uncertainties." Consequently, we identify that cost growth is not a single static number but a range of values, and recognize there is a possibility that costs could go up or down in price. Thus, an element of risk is involved in cost growth. This point may seem obvious but it is crucial to understanding the characteristics of risk and uncertainty. Our research begins with the knowledge that

cost growth encompasses elements of both uncertainty and risk. Uncertainty implies an alternate value(s) can exist and risk is the chance of incurring a gain or loss as a result of the alternate value(s).

Characteristics

The Air Force Materiel Command's (AFMC) *Financial Management Handbook*, clearly states, "cost estimating deals with uncertainty." The dilemma is that cost estimates try to calculate the cost of a system that will be designed, constructed, and completed in the future. It is the cost estimator's job to quantify the possible probability distribution associated with that future cost (*AFMC Financial Management Handbook*, 2001:11). The cost estimate is simply one value or one prediction of that event. The AFMC Handbook describes 'risk' as the effect from uncertainties and consequences of future events, and "risk is the summation of the probable effect of unknown elements in technical, schedule or cost related activities within a program" (*AFMC Financial Management Handbook*, 2001:11). The wording of this definition, suggests some type of valuation, in terms of dollars, be made for these separate areas along with a probability distribution to represent the associated range of possible values. Consequently, our research quantifies risk as the unknowns in terms of the characteristics of technical, schedule and cost, and also includes a probability distribution to show the range of values.

Risk Estimating Methods

We now focus on methodologies used to assess probabilistic values. Within the cost estimating community several methods exist to assess and quantify risk. Each

method's use depends on many factors including: type of estimate, type of risk, estimate accuracy, level of detail needed, estimator skill, and time to complete the estimate.

The AFMC handbook details three methods for assessing the likelihood of an event occurring: a posteriori, a priori, and subjective judgment:

> 1) The first method, a posteriori, or "after the fact" relationship to past events (direct knowledge), is based on some previous occurrence such as the cost outcome of previous projects conducted by the organization. If enough samples from the past history (the population) are drawn, the probability of the next event occurring in a particular way may be estimated. A methodology like Monte Carlo simulation may also be used. The Monte Carlo simulation is conducted where the analyst determines the probability of future events by using an experimental model to approximate expected actual conditions. Such a model is fashioned from previous histories of similar projects.
> 2) Sometimes a distribution of possible outcomes for an event is not based on experience or sampling but on a priori, or "before the fact" theoretical probability distribution. The use of the closeness of the assumptions used in developing the theoretical distribution is to the real world situation being analyzed.
> 3) Many times an analyst will have to use a subjective judgment (indirect knowledge) in estimating probability. This approach relies on the experience and judgment of one or more people to create the estimated probability distribution. The result is known as a subjective probability. A distribution estimate is an analysis by one or more informed persons of the relative likelihood of particular outcomes of an event occurring. Distribution estimates are subjective. An example of this approach is the Delphi method. (*AFMC Financial Management Handbook*, 2001:8-9; Sipple, 2002:14-15)

The Ballistic Missile Defense Organization (BMDO) cost estimating community utilizes a spectrum of five different risk assessment techniques to prepare estimates. The application of the five methods differs by the degree of difficulty and the required precision (accuracy) needed in the estimate. Figure 2 shows a chart of BMDO's risk methods (Coleman, 2000:4; Sipple, 2002:17).

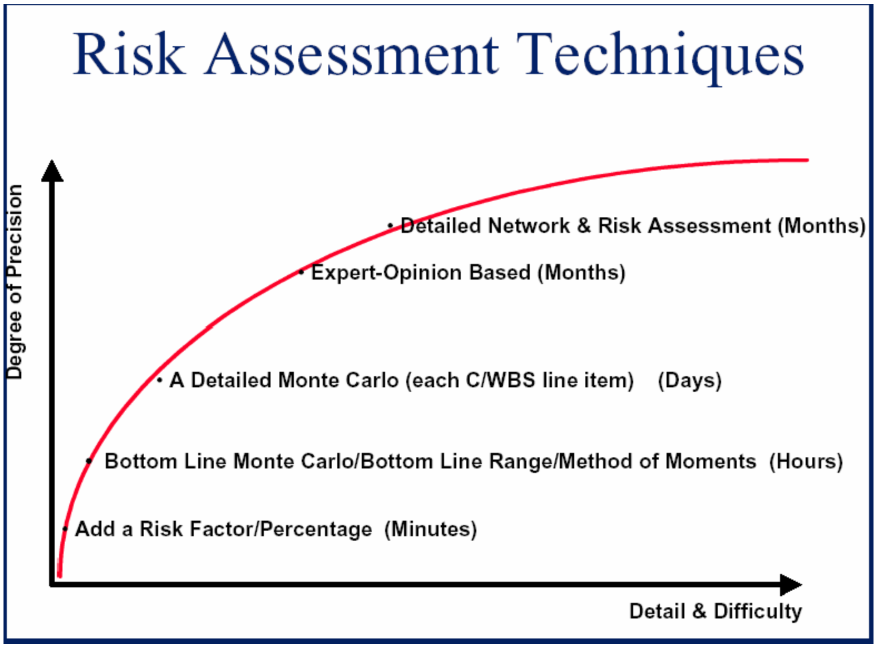


Figure 2 - Risk Assessment Techniques (Coleman, 2000:4-9)

A brief explanation of each of the methods in Figure 2 is detailed below:

- Detailed Network and Risk Assessment: is the most precise and most difficult to apply. It requires a very detailed schedule and task breakout. It uses a beta or triangular distribution to schedule item durations and creates a stochastic model from which to estimate the risk of a schedule slip. The estimator uses the Monte Carlo Simulation method to estimate the cost (Coleman, 2000:4-9).
- Expert-Opinion-Based: relies on surveys of experts to determine the possible distributions of Work Breakdown Structure (WBS) item costs. Uses Monte Carlo simulation to estimate a range of possible costs. Assumes experts are accurate (Coleman, 2000:12).

- Detailed Monte Carlo Simulation: C/WBS is the Cost or Work Breakdown Structure. Uses Monte Carlo Simulation, but relies on historical data to develop probability distributions of cost outcomes (Coleman, 2000:16).
- Bottom Line Monte Carlo/Bottom Line Range/Method of Moments: may use Monte Carlo Simulation, but on higher levels of the WBS. Other uses include a limited database, analogy methodology or expert opinion to determine risk estimates (Coleman, 2000:4).
- Add a Risk Factor/Percentage: is the least precise and easiest technique to use. Relies on technical expert judgment to assign a high-level, subjective risk factor for the estimate (Coleman, 2000:4).

Past Research in Cost Growth

Our goal is to "realistically estimate" costs and ultimately build a prediction model for cost growth within the EMD phase of acquisition. We have looked at what cost growth is, how it is calculated and the environmental factors that influence it. We now turn our attention to past research efforts in seeking further insight into the causes of cost growth.

Much has been written in the past regarding cost growth analysis. For example, in James A. Gordon's 1996 thesis, he complies a partial historical listing of studies conducted on the subject by RAND and AFIT (see Tables 1 and 2). While each of these studies provides valuable clues to understanding the characteristics and causes of cost growth, each also differs from the study at hand in purpose, scope, or methodology. We find one study that uniquely encapsulates much of the previous cost growth research and

applies it to a scope similar to ours: Vincent Sipple's (2002) thesis. Hence, we utilize Sipple's (2002) thesis for its exhaustive research, meticulous detail, and correlation with our study as a benchmark for our research effort.

Table 1 - RAND Reports (Gordon, 1996:2-2)

Author (Year)	Findings	Sensitivity Factors
Jarvaise, et al. (1996)	Defense System Cost Performance Database	Derived from SARs
Drezner, et al. (1993)	Cost Estimates biased toward underestimation by about 20% from PE and DE and 2% from PdE	Program Size, Maturity
Drezner (1992)	No demonstrated relationship between prototyping and cost or schedule outcomes (67)	No Program Phase, Not System Type
Hough (1992)	Selected Acquisition Reports can Delay, Mask or Exclude Significant Cost Growth	Economic, Quantity, Schedule, Engineering, Estimating and Other Changes

Table 2 - AFIT Theses (Gordon, 1996:2-3)

Author (Year)	Findings	Sensitivity Factors
Nystrom (1996)	Complex non-linear EAC methods not superior to simpler index based EAC methods	Stage of Completion, System Type, Program Phase, Contract Type, Service Component, and Inflation
Buchfeller and Kehl (1994)	No Significant Differences in Cost Variances between categories	Not Service, Not Program Phase, Not Contract Type, Not Stage of Completion
Elkinton and Gondeck (1994)	BAC Adjustment Factors derived from Historical "Cost Growth" do not Improve EACs	Not Contract Type, Not Stage of Completion
Pletcher and Young (1994)	Contracts which Improved Cost Performance over time differ from those which Worsen	Performance Management Baseline Stability
Terry and Vanderburgh (1993)	SCI based EAC best predictor of CAC for all Stages of Contract Completion	Contract Completion Stage, Program Phase, Contract Type, Service Component, System Type, Major Baseline Changes, but not Management Reserve
Wandland (1993)	Completed Contracts have more "Cost Growth" than Sole Source	Not Contract Type, Not Absolute Price
Wilson (1992)	Cost Overruns at Completion are Worse than between 15 and 85% complete (α =.15)	Service (except Navy), Contract Type, System Type, and Program Phase, but not relative time
Singleton (1991)	"Cost Growth" can be predicted based on three factors	Schedule Risk, Technical Risk and Configuration Stability
Obringer (1988)	"Cost Growth" is not attributable to increased Industry Direct or Overhead to Total Cost Ratio	Specific Contractors (8 of 16) showed growth between 1980 and 1986
Blacken (1986)	"Cost Growth" varies with Characteristics of Contract Changes	Scope, Number of Effected SOW Pages, Contract Type, Change Type, Time to Definitize, Time to Negotiate, Not to Exceed Estimate, Stage of Completion, Stage of Development, Schedule Changes, Length of ECP, Length of Period of Performance

Sipple (2002) provides a comprehensive review of the 12 previous cost growth studies listed in Table 3. Sipple extracts numerous bits of data for developing predictor variables from each of these studies, as well as, valuable insight to the root causes of cost

growth. Of these, we take particular note of the NAVAIR and 1993 RAND findings as these studies most closely align with our research.

Table 3 - Sipple Thesis (Sipple, 2002:20-44)

Author (Year)
IDA (1974)
Woodward (1983)
Obringer (1988)
Singleton (1991)
Wilson (1992)
RAND (1993)
Terry & Vanderburgh (1993)
BMDO (2000)
Christensen & Templin (2000)
Eskew (2000)
NAVAIR (2001)
RAND (2001)

The NAVAIR, study is significant to us because it evaluates cost growth, from SAR data (our database) through the implementation of "cohort tracking" (Dameron, 2001:7). The term "cohort tracking" is used to group cost growth according to similar characteristics. The five groups they identify are:

- RDT&E cost growth for programs with a planning estimate (PE) and a development estimate (DE),
- RDT&E cost growth for programs with a DE only,
- Procurement cost growth for programs with a PE, a DE, and a production estimate (PdE),
- Procurement cost growth for programs with a DE and a PdE only,
- Procurement cost growth for programs with a DE only (Dameron, 2001:10).

Thus, the use of cohort tracking isolates what we seek to predict; RDT&E cost growth with a DE estimate. Specifically, NAVAIR finds the PE and DE cohort has 30 percent RDT&E cost growth and the DE-only cohort has 25 percent RDT&E cost growth. NAVAIR also finds a significant linkage between the phases of acquisition and cost growth and between the appropriations. Of particular interest to us, is a strong connection between RDT&E cost growth in the PDRR phase and RDT&E in the EMD phase (Dameron, 2001:14). Such knowledge, offers us a "leading" indicator for EMD RDT&E cost growth. Additionally, NAVAIR finds a link between cost growth in the RDT&E appropriation and the procurement appropriation.

These findings indicate a substantial "forward" roll of costs as a program develops overtime. Thus, such findings corroborate the historical cost growth trend cited by Drezner and drive home the need for research in this area. We take away the knowledge that if cost growth appears at any phase of development, subsequent phases will also experience cost growth. Such insight leads us to consider some type of leading indicator in our models to forecast cost growth, as well as, opens the door for possible follow-on research to connect EMD to the production phase and the PDRR phase to EMD.

The RAND 1993 study is noteworthy due to its use of (and extensive history with) the SAR data and its prominence in the cost growth analysis arena. Within DoD, RAND methodologies and practices are usually the de-facto standard. RAND establishes that inflation and quantity have the greatest effect on cost growth. Yet, since these two factors are already included as a basic premise of a cost estimate, RAND establishes a procedure of excluding them from their data before analyzing cost growth. To be

consistent, we also will follow this approach in our research. RAND enumerates on several other factors that relate to cost growth but ultimately concludes, “no single factor explains a large portion of the observed variance in cost growth outcomes” (Drezner, 1993: 49).

Sipple argues the reason RAND draws this conclusion is that it “comes from a top-level, exploratory analysis of the total cost growth data. Whereas, RAND finds no significant explanatory variables for overall cost variance, the possibility exists that breaking down cost growth into its components might uncover some significant explanatory variable” (Sipple, 2002:35). For the purpose of this research, we utilize the predictor variables detailed by RAND but follow Sipple’s methodology of using those variables to predict cost growth in single compartmentalized area vice an overall approach.

Sipple’s (2002) work seeks to predict cost growth of RDT&E accounts in the EMD phase of acquisition, using a SAR database spanning 1990 – 2000. Sipple measures cost growth as a percentage increase in cost from the DE, as recorded in the SAR, and focuses specifically on predicting the SAR cost growth category of “Engineering.” Sipple first identifies the existence of a mixture distribution – a discrete point mass coupled with a continuous distribution. In this case, the discrete mass, centered on zero, represents programs with zero cost growth.

To account for the mixture distribution, Sipple uses a unique and innovative (for the cost community) two-step process to estimate cost growth. The two-step process entails the use of first, logistic regression to distinguish between those programs that have

cost growth and those that do not. Second, the use of multiple regression to predict the amount of cost growth that will occur given there is cost growth.

To the best of our knowledge, Sipple is the first to use logistic regression for cost estimation purposes. Logistic regression predicts a binary (1/0 or yes/no) response from discrete data. Sipple demonstrates through the use of four regression models (A, B, C, D) that the combination of logistic and multiple regression produce similar predictive results as a traditional single-step multiple regression cost estimating methodology. However, the two-step methodology is preferred to the single-step methodology because of the stronger statistical foundation achieved with the two-step method.

First, Sipple builds model A to predict whether a *program will have cost growth (yes or no?)* using logistic regression. Model A uses all 90 data points in Sipple's database which represent programs with both positive and negative cost growth. Programs with positive cost growth are converted to a "yes" response while zero or negative cost growth programs are converted to a "no" response. Next, model B is built to predict the amount of cost growth that will occur using only those programs that experience cost growth (47 of the 90 data points have cost growth). A log transformation of the Y response is used to correct for heteroskedasicity or non-constant variance of the residuals. Sipple finds that without such a transformation none of the models pass the underlying Ordinary Least Squares (OLS) statistical assumptions test of normality and constant variance. The use of models A and B together is then established as the "two-step" process baseline for comparison with the other models.

Model C is built as an alternative to model B except that the Y response is not transformed. Hence, model C is built from the same 47 data points as model B but does

not correct for the statistical assumptions tests. Sipple uses this model to compare the difference in predictive ability of a model without statistical foundation to the B model with correct statistical assumptions. Accordingly, Sipple finds that none of these C models pass the tests for normality and constant variance. Lastly, model D is built to ascertain the effects of not recognizing the mixture distribution and overlooking the OLS statistical assumptions. Thus, model D is created using the entire 90 data point set (without logistic regression) and the *Y* response is not transformed. This model tests the effects of the traditional single-step approach to cost estimating versus the two-step (model A & B) combination. Sipple uses stepwise regression to build this model and again ignores OLS statistical assumptions tests (because all models fail without the log transformation).

Sipple validates all four models with a 25-point "withhold" data set which represents 20 percent of his initial data set. Of the 25 data points, 12 have missing values for model A leaving 13 data points for validation of this model. Sipple demonstrates that a seven-variable logistic regression model (A) accurately predicts 9 out of 13 data points during validation for an almost 70 percent accuracy rate (Sipple, 2002, 82). For, model B's validation, only 14 of the 25 data points are used (11 have no cost growth). Sipple finds a three variable OLS model is the preferred model, with an Adj R^2 0.4645 and validates with 69.23 percent of observations within an 80 percent prediction bound (Sipple, 2002:87-88).

For model C, Sipple finds that models B and C perform "on par" with each other except that doubts about model C's inferential uncertainty overshadow the results. Model C's non-transformed *Y* response precludes it from passing the Shapiro-Wilk test

for normality and the Breush-Pagan test for constant variance. Furthermore, significant influential outliers exist which could not be eliminated from the data without causing more data points to become influential (Sipple, 2002:90). For model D's validation, the entire 25 point withhold data set is used to mirror the premise of model D (i.e., single step cost estimation). Sipple finds model D's results similar to that of model C's – failure of any of the models to pass statistical assumptions tests for normality and constant variance of the residuals, as well as, the existence of numerous influential data points. Hence, the results from models C and D are unreliable and dubious at best for drawing statistical conclusions (Sipple, 2002:117).

Our research seeks to expand Sipple's findings in that we seek to predict cost growth in the SAR categories of Estimating, Schedule, Support and Other, using the methodology of models A and B only. Models C and D are not duplicated since these models use are not reliable for cost estimators.

Chapter Summary

In this chapter, we outline an operational understanding and knowledge of what cost growth means, how it is calculated and the genetic make-up of DoD cost growth. We reference past cost growth studies regarding the causes of cost growth and obtain clues of possible predictor variables to use in our research. We follow this literature review in the next chapter by highlighting our methodology to build upon Sipple's (2002) work.

III. Methodology

Chapter Overview

This chapter enumerates the procedures we use to perform this research. We first discuss our database and its limitations. We follow with details of our data collection process and list candidate variables for model development. Finally, we discuss our exploratory data analysis results and our methodology for performing logistic and multiple regression.

Database

The Selected Acquisition Reports (SAR) are the source data for this study. The SARs contain a plethora of programmatic information on each major acquisition program of the DoD. Each program included in the SAR submits specific required information annually to SAR administrators, currently the Office of the Under Secretary of Defense for Acquisition and Technology. This information is categorized into nineteen different sections of the SAR and includes historical, schedule, cost, budget, and performance information. The SAR only reports on programs that meet specific dollar thresholds, which constitute DoD's most visible and highest interest level programs, otherwise known as ACAT IC or D programs (Knoche, 2002:1).

Although, the specific ACAT reporting criteria changes over time, the SAR database consistently represents programs that are the U.S government's most vital. As such, the majority of the programs included in the SAR carry some level of security classification: classified, confidential or restricted. For our research, we collect only limited programmatic and cost data, which is normally not classified. However, if the

information we seek on an individual program is specifically classified, we omit the use of that piece information in our research.

The SAR format provides two estimates for each program: the baseline estimate and the current estimate. The SAR may also include a third estimate, one of the overall "approved program" which reflects the latest program decision memorandum (Hough, 1992:4). The SAR catalogs any deviations from these "programmed budgets" into one of seven different cost variance categories. The cost variances are reported in both base-year (year of initial program funding) and then-year (base-year adjusted for inflation) dollars. A program's total cost variance is then the sum of these seven cost variances. The seven SAR cost variance categories are:

- Economic: changes in price levels due to the state of the national economy
- Quantity: changes in the number of units procured
- Estimating: changes due to refinement of estimates
- Engineering: changes due to physical alteration
- Schedule: changes due to program slip/acceleration
- Support: changes associated with support equipment
- Other: changes due to unforeseen events (Hough, 1992:5; Drezner, 1993:7)

Our research uses the base-year dollar cost variances to conduct data analysis. We choose base-year dollars, which exclude inflationary affects, so that we can easily convert individual estimates into a single base year and then draw comparisons between programs. We convert all program estimates to CY $2002 dollars so that we can evaluate

cost growth in terms of today's dollars. Additionally, we focus only on programs that have a Development Estimate (DE) as their baseline estimate as reported in the SAR.

By convention, when analyzing cost growth, the cost analyst routinely normalize the data to account for the effects of inflation and quantity changes, since these items can have a substantial impact on overall cost growth. Our research also follows this convention but we do not make manual adjustments to the program data, since the SAR pre-computes these values and incorporates them as two of the seven cost variance categories (quantity and economic).

As mentioned in chapter II, we follow many of the procedures laid out in the 1993 RAND report yet, in one situation we diverge. RAND utilizes only positive cost variances (growth) in its analysis. In contrast, our study takes into account both zero and negative cost variances for use in our logistic regression analysis and model building. Thus, we collect all cost variance data, not just exclusively positive variance.

Also discussed in chapter II, an area of consternation in computing cost growth is the identification of which baseline to best measure cost growth from. The SAR offers three different possible baseline estimates from which to choose; the planning estimate (PE), the development estimate (DE), and the production estimate (PdE). These estimates occur before the start of Milestone I, II, and III, respectively. According to RAND, cost estimates performed later in a programs life cycle are more accurate and reflect improved program information and reduced risk. This is logical since program uncertainty (risk) equates to greater variation in cost estimates, and as uncertainty is reduced, cost estimates (accuracy) improve. Thus, it follows that cost growth increases as the baseline used to measure cost growth moves back time (Hough, 1992: 10-11). For

our research, we are concerned only with the cost growth in RDT&E accounts during EMD. Thus, we choose to use only programs with a DE baseline estimate to capture the cost growth during the entire EMD phase. (See Figure 1 in chapter II for a reference of the acquisition timeline.)

According to RAND, cost growth is defined as "the difference between the most recent or final estimate of the total acquisition cost for a program and the initial estimate" (Hough, 1992:10). The first or initial estimate can be a PE, DE, or PdE depending on the program. Our research uses only programs with a DE baseline estimate as the initial estimate since we focus on cost growth in the EMD phase. We compute cost growth as a percentage (this is explained in more detail later in this chapter) cost growth by first calculating the difference of the current estimate minus the DE. We then divide the result by the DE. Fortunately, the SAR data contains all the necessary information to make these calculations and supports our methodologies.

SAR Limitations

Although the SAR is the primary source of research into cost growth, its use is not without limitations. In the 1992 RAND report by Paul Hough, he notes that while the government has implemented many reporting changes that continually improve the "quality and comprehensiveness of the data," the SAR still possesses numerous difficulties with respect to cost growth calculations. According to RAND, these problems include:

- Failure of some programs to use a consistent baseline cost estimate
- Exclusion of some significant elements of cost

- Exclusion of certain classes of major programs (e.g., special access programs)
- Constantly changing preparation guidelines
- Inconsistent interpretation of preparation guidelines across programs
- Unknown and variable funding levels for program risk
- Cost sharing in joint programs
- Reporting of effects of cost changes rather than their root causes (Hough, 1992:v; Sipple, 2002:49)

Most literature agrees that the SAR provides some consistency in the reporting of program data, however, interpretations of the specific reporting guidelines vary from program to program, which increases inconsistency of reporting. Additionally, the specific reporting guidelines themselves change over time, further adding to the inconsistency of the data (Hough, 1992:4). Notwithstanding the noted data limitations, RAND recognizes the SAR as "the logical source of data for calculating cost growth on major procurements" (Hough, 1992:9). Thus, our study follows RAND's lead and adopts the SAR as our source of program data from which to estimate cost growth.

The Baseline Problem

Once a cost growth baseline is selected the analyst must recognize that the "selected" baseline may not be consistent over time or from program to program. This inconsistency stems from two types of events: rebaselining and evolutionary changes. Rebaselining occurs when the program office develops a new baseline estimate in the middle of an acquisition phase. The new program estimate replaces the old estimate; yet, it retains the original estimate's designation (PE, DE, or PdE). Evolutionary model

changes occur when modifications are made to a program such that the "current model only remotely resemble what was originally estimated" (Hough, 1992:12-14). Detecting either a rebaselined or evolutionary changed program from a non-changed program is difficult at best and extremely hard to normalize out of SAR data (Hough, 1992:12-14).

Variation of Reported Program Costs

Congress continuously changes the SAR preparation guidelines in an effort to improve quality. While these changes usually have no direct monetary impact on the program, they do present problems of accuracy and consistency for the cost growth analyst. Variation in reporting requirements makes accurate calculation of cost growth difficult (Hough, 1992:12-47). Moreover, RAND describes the practice of postponing the reporting of cost growth as a more systemic problem. Postponement occurs when program managers do not report cost growth until after a significant milestone decision has passed, presumably to appear "lower" cost. Thus, cost growth is erroneously allocated to the incorrect program phase, further exploiting the difficulty in accurate cost growth analysis.

Inconsistency in SAR Preparation Guidelines and Techniques

Closely associated with the problems of changing reporting requirements is the problem of inconsistent application of these changes. While changes arguably improve the overall SAR content quality, the consistency and uniformity of the data is tainted over time. Such fluctuations in the database make program comparisons difficult. Magnifying this problem is that not all organizations interpret and adopt changes at the same time. RAND acknowledges that, "after a major change, consistency among SARs is not

ensured until all programs with current reporting use the same set of rules" (Hough, 1992:19-20).

Incomplete Database

According to Hough, when analyzing cost growth, care should be taken to ensure the sample size of data used is representative of the overall population and that "...quality studies on cost growth should identify what portion of the total [SAR] population is included and why the sample is representative of the whole or is satisfactory for meeting the study objectives." Unfortunately, the SAR database is incomplete to start with, since it does not include lower dollar value (below ACAT 1D) DoD programs, or "highly sensitive – classified" or "black" programs (Hough, 1992:17). According to the SAR instructions, any programs deemed by the Secretary of Defense to be "highly sensitive – classified" are exempt from SAR reporting. By some estimates, the percentage of "black" programs represents, at least, 20 percent of the DoD acquisition budget (Hough, 1992:17). Thus, SAR based cost growth research includes only a portion of the total DoD pool of acquisition programs in existence.

Unknown Funding Levels for Programs

Maintaining key program funding with a declining DoD budget, makes program funding less stable. As a result, Congress and the services often take money from one program to fund another. To counteract this, program managers and cost estimators often include a cushion or monetary padding to account for this risk in their estimates. This cushion, known officially as management reserve funding, is often hidden among one or

more budget line items. Thus, the SAR may already reflect some estimate for risk, however, identifying these risk dollars is virtually impossible.

Joint Programs

Some major weapons programs are developed and used by more than one service component. This leads to uniformity problems within the SAR. In joint programs, investment costs can be equally distributed among all the participants, borne entirely by one component, or allocated on some other percentage distribution. No guidelines exist to govern such programs or allocations. Consequently, no single methodology is used within the SAR database, which further adds to the inconsistency of the database.

Reporting Effects of Cost Changes Rather Than Root Causes

RAND recognizes that current SAR requirements do not disclose the "root causes" of cost growth. The SAR reports seven different categories of cost variance for each program but does not specifically report on what actually drives a program's cost. Although a thorough review of other SAR sections might give an indication of the "root cause" of cost growth, there is no guarantee of this happening. Hence, this limitation hampers the cost analyst's quest for the true drivers of cost growth (Hough, 1992:23).

Although, RAND openly acknowledges the many limitations of the SAR database, these limitations do not deter its use for analyzing cost growth. A SAR database has many advantages including: strict reporting format (which improves consistency of the data), annual SAR training for those submitting SAR reports (which also improves consistency of the data (Knoche, 2002:2.B.3.2)), and increased scrutiny of data (because SARs go before Congress, the data is more reliable). Thus, we

acknowledge that all sources of cost growth data contain some reporting, format, or other inaccuracies however; SAR data has its benefits and is widely recognized as the best option available for cost growth analysis. Hence, we adopt the SAR as our database for this research study.

Data Collection

Since our research seeks to build upon Sipple's (2002) work, we start with his established SAR database. Sipple's database includes RDT&E and procurement program data collected from SAR reports of programs that use the DE as its baseline estimate (Sipple, 2002:57). Sipple systematically collected individual program data beginning with the December 2000 SAR and worked backwards in time to 1990, collecting sufficient data to support a statistically significant regression. Furthermore, only one SAR for each program (the latest) was included to ensure independence of the data points (Sipple, 2002:57). In many instances, he notes that the most recent DE based SAR for a program is the last SAR of the EMD phase of acquisition for that program, or it may be the last reported SAR due to program completion or termination. As discussed earlier, he excludes those SAR programs that contain sensitive information or which are restricted with a security classification.

We start our data collection with a thorough review of the most recently released SAR, specifically December 2001. This SAR represents the next successive SAR from where Sipple ends his data collection. Inclusion of this SAR information extends our research database to include RTD&E and procurement programs using a DE estimate and having an EMD phase of development from 1990 to 2001.

We begin by updating the current estimates of any programs presently included in our database. That is, we ensure our database of SAR programs utilizes the most recently available program data. We then add the programmatic data of any new programs, which meet our research criteria of RDT&E programs using a DE as their baseline estimate that are not currently included in the database. We include all programs that meet this criterion. We do not exclude joint service programs simply because of the previously identified inconsistency in reporting investment allocation costs between multiple program beneficiaries. Further, to maintain consistency with Sipple's (2002) work we do not collect or use classified SAR program data. Lastly, the specific type of program data we extract from the SAR mirrors Sipple's original methodology, except that we use this information in predicting cost growth in four separate SAR categories (Estimating, Schedule, Support and Other) versus a single area.

Exploratory Data Analysis

Sipple (2002) found the data used for analysis possessed a mixture distribution. Consequently, we also encounter a mixture distribution within our data set. A mixture distribution refers to a response variable whose data comprises of continuous and discrete data. For our study, the discrete data centers at zero, i.e., no cost growth. Using statistical analysis methods, the general solution to a mixture distribution calls for splitting the data into two separate sets, one for continuous and the other for discrete data. This is required because the probability of obtaining a specific number within a continuous distribution is zero, which no longer holds for a discrete mass.

The mixture distribution dictates that we use a two-step methodology in order to analyze using statistical methods. The first step, utilizes logistic regression to analyze the discrete data. The second step, utilizes multiple regression to analyze the continuous data. Hence, we develop two types of models for our research objective. Model A, for logistic regression to predict whether or not a program will have cost growth from the full data set, and model B for multiple regression to predict the amount of cost growth from only those programs that experience cost growth (Sipple, 2002:58-59).

Upon further evaluation of the data, we also observe that several programs have negative SAR cost variances. We speculate that negative values normally do not occur since a cost estimator would never assign such a value to cost estimate. However, for our logistic regression model we consider all values, negative or positive. To do this, we simply convert all negative cost growth figures to zero for inclusion in our logistic regression model.

Finally, before starting the actual analysis of our data we set aside 20 percent of our data for validation purposes. We sequentially input all the program data (# 1-122) into our statistical software program JMP® 4.0 (SAS Institute, 2001), and then utilize the random shuffle feature within JMP® to independently randomize the data. We then remove the top 25 rows of randomized data, which corresponds to 20 percent of the entire database, for use in validating our models later. We do not use this data during the model building process.

Response Variables

Our research focus is to locate predictors of cost growth due to Estimating, Schedule, Support, and Other changes within the RDT&E accounts. The SAR report identifies these categories as cost variances for both the RDT&E and the procurement appropriations. However, we limit our focus to only the RDT&E accounts only. Since we have a mixture distribution, we use two different response variables. One variable indicates if cost growth will occur while the second variable conveys the magnitude of cost growth. We express the first variable as a binary variable where a '1' means that we estimate a program will experience cost growth, while a '0' means it will not. We call this variable *R&D Cost Growth?* (Sipple, 2002:60).

We choose the second variable to have the form of a percentage, rather than a dollar amount to apply equally to both large and small programs. We prefer the percentage-based variable to the dollar-based variable since it eliminates the need to quantify between programs of different sizes. In essence, it equalizes programs of different sizes for comparison purposes. Thus, we focus on predicting the percentage change in RDT&E cost growth due to Schedule, Estimating, Support, and Other changes in our models. We call the response variables: *Schedule %, Estimating%, Support% and Other%.*

Predictor Variables

Our research uses the pool of candidate variables amassed by Sipple (2002). The variables, all derived from literature review sources, are proven predictors of cost growth.

Thus, we use these predictor variables in our quest to build a tool for cost estimators that accurately predicts EMD cost growth for RDT&E accounts.

Sipple groups the predictor variables into five broad categories: program size, physical type of program, management characteristics, schedule characteristics, and other characteristics. Within these broad categories, he also creates several subcategories levels to further group similar variables. For example, the physical type category is further divided into 'domain of operation variables' and 'functional variables' (Sipple, 2002:61). We modify one predictor from Sipple's original definition and rename it: New Concurrency Measure % to reflect a computational change. Listed below are the predictor variables sorted by category and subcategories. Short descriptions are provided for clarity:

Program Size Variables

- *Total Cost CY $M 2002* – continuous variable which indicates the total cost of the program in CY $M 2002
- *Total Quantity* – continuous variable which indicates the total quantity of the program at the time of the SAR date; if no quantity is specified, we assume a quantity of one (or another appropriate number) unless the program was terminated
- *Prog Acq Unit Cost* – continuous variable that equals the quotient of the total cost and total quantity variables above
- *Qty during PE* – continuous variable that indicates the quantity that was estimated in the Planning Estimate
- *Qty planned for R&D$* – continuous variable which indicates the quantity in the baseline estimate

Physical Type of Program

- Domain of Operation Variables
    - *Air* – binary variable: 1 for yes and 0 for no; includes programs that primarily operate in the air; includes air-launched tactical missiles and strategic ground-launched or ship-launched missiles
    - *Land* – binary variable: 1 for yes and 0 for no; includes tactical ground-launched missiles; does not include strategic ground-launched missiles

  - *Space* – binary variable: 1 for yes and 0 for no; includes satellite programs and launch vehicle programs
  - *Sea* – binary variable: 1 for yes and 0 for no; includes ships and ship-borne systems other than aircraft and strategic missiles
- Function Variables
  - *Electronic* – binary variable: 1 for yes and 0 for no; includes all computer programs, communication programs, electronic warfare programs that do not fit into the other categories
  - *Helo* – binary variable: 1 for yes and 0 for no; helicopters; includes V-22 Osprey
  - *Missile* – binary variable: 1 for yes and 0 for no; includes all missiles
  - *Aircraft* – binary variable: 1 for yes and 0 for no; does not include helicopters
  - *Munition* – binary variable: 1 for yes and 0 for no
  - *Land Vehicle* – binary variable: 1 for yes and 0 for no
  - *Ship* – binary variable: 1 for yes and 0 for no; includes all watercraft
  - *Other* – binary variable: 1 for yes and 0 for no; any program that does not fit into one of the other function variables

Management Characteristics

- Military Service Management
  - *Svs > 1* – binary variable: 1 for yes and 0 for no; number of services involved at the date of the SAR
  - *Svs > 2* – binary variable: 1 for yes and 0 for no; number of services involved at the date of the SAR
  - *Svs > 3* – binary variable: 1 for yes and 0 for no; number of services involved at the date of the SAR
  - *Service = Navy Only* – binary variable: 1 for yes and 0 for no
  - *Service = Joint* – binary variable: 1 for yes and 0 for no
  - *Service = Army Only* – binary variable: 1 for yes and 0 for no
  - *Service = AF Only* – binary variable: 1 for yes and 0 for no
  - *Lead Svc = Army* – binary variable: 1 for yes and 0 for no
  - *Lead Svc = Navy* – binary variable: 1 for yes and 0 for no
  - *Lead Svc = DoD* – binary variable: 1 for yes and 0 for no
  - *Lead Svc = AF* – binary variable: 1 for yes and 0 for no
  - *AF Involvement* – binary variable: 1 for yes and 0 for no
  - *N Involvement* – binary variable: 1 for yes and 0 for no
  - *MC Involvement* – binary variable: 1 for yes and 0 for no
  - *AR Involvement* – binary variable: 1 for yes and 0 for no
- Contractor Characteristics
  - *Lockheed-Martin* – binary variable: 1 for yes and 0 for no
  - *Northrup Grumman* – binary variable: 1 for yes and 0 for no
  - *Boeing* – binary variable: 1 for yes and 0 for no
  - *Raytheon* – binary variable: 1 for yes and 0 for no

- *Litton* – binary variable: 1 for yes and 0 for no
- *General Dynamics* – binary variable: 1 for yes and 0 for no
- *No Major Defense KTR* – binary variable: 1 for yes and 0 for no; a program that does not use one of the contractors mentioned immediately above = 1
- *More than 1 Major Defense KTR* – binary variable: 1 for yes and 0 for no; a program that includes more than one of the contractors listed above = 1
- *Fixed-Price EMD Contract* – binary variable: 1 for yes and 0 for no

Schedule Characteristics

- RDT&E and Procurement Maturity Measures
    - *Maturity (Funding Yrs complete)* – continuous variable which indicates the total number of years completed for which the program had RDT&E or procurement funding budgeted
    - *Funding YR Total Program Length* – continuous variable which indicates the total number of years for which the program has either RDT&E funding or procurement funding budgeted
    - *Funding Yrs of R&D Completed* – continuous variable which indicates the number of years completed for which the program had RDT&E funding budgeted
    - *Funding Yrs of Prod Completed* – continuous variable which indicates the number of years completed for which the program had procurement funding budgeted
    - *Length of Prod in Funding Yrs* – continuous variable which indicates the number of years for which the program has procurement funding budgeted
    - *Length of R&D in Funding Yrs* – continuous variable which indicates the number of years for which the program has RDT&E funding budgeted
    - *R&D Funding Yr Maturity %* – continuous variable which equals *Funding Yrs of R&D Completed* divided by *Length of R&D in Funding Yrs*
    - *Proc Funding Yr Maturity %* – continuous variable which equals *Funding Yrs of R&D Completed* divided by *Length of Prod in Funding Yrs*
    - *Total Funding Yr Maturity %* – continuous variable which equals *Maturity (Funding Yrs complete)* divided by *Funding YR Total Program Length*
- EMD Maturity Measures
    - *Maturity from MS II in mos* – continuous variable calculated by subtracting the earliest MS II date indicated from the date of the SAR
    - *Actual Length of EMD (MS III-MS II in mos)* – continuous variable calculated by subtracting the earliest MS II date from the latest MS III date indicated
    - *MS III-based Maturity of EMD %* – continuous variable calculated by dividing *Maturity from MS II in mos* by *Actual Length of EMD (MS III-MS II in mos)*
    - *Actual Length of EMD using IOC-MS II in mos* – continuous variable calculated by subtracting the earliest MS II date from the IOC date

- *IOC-based Maturity of EMD %* – continuous variable calculated by dividing *Maturity from MS II in mos* by *Actual Length of EMD using IOC-MS II in mos*
- *Actual Length of EMD using FUE-MS II in mos* – continuous variable calculated by subtracting the earliest MS II date from the FUE date
- *FUE-based Maturity of EMD %* – continuous variable calculated by dividing *Maturity from MS II in mos* by *Actual Length of EMD using FUE-MS II in mos*

- Concurrency Indicators
  - *MS III Complete* – binary variable: 1 for yes and 0 for no
  - *Proc Started based on Funding Yrs* – binary variable: 1 for yes and 0 for no; if procurement funding is budgeted in the year of the SAR or before, then = 1
  - *Proc Funding before MS III* – binary variable: 1 for yes and 0 for no
  - *Concurrency Measure Interval* – continuous variable which measures the amount of testing still occurring during the production phase in months; actual IOT&E completion minus MS IIIA (Jarvaise, 1996:26)
  - *New Concurrency Measure %* – continuous variable which measures the percent of testing still occurring during the production phase; (MS IIIA minus actual IOT&E completion in moths) divided by (actual minus planned IOT&E dates) (Jarvaise, 1996:26)

Other Characteristics

- *# Product Variants in this SAR* – continuous variable which indicates the number of versions included in the EMD effort that the current SAR addresses
- *Class – S* – binary variable: 1 for yes and 0 for no; security classification Secret
- *Class – C* – binary variable: 1 for yes and 0 for no; security classification Confidential
- *Class – U* – binary variable: 1 for yes and 0 for no; security classification Unclassified
- *Class at Least S* – binary variable: 1 for yes and 0 for no; security classification is Secret or higher
- *Risk Mitigation* – binary variable: 1 for yes and 0 for no; indicates whether there was a version previous to SAR or significant pre-EMD activities
- *Versions Previous to SAR* – binary variable: 1 for yes and 0 for no; indicates whether there was a significant, relevant effort prior to the DE; a pre-EMD prototype or a previous version of the system would apply
- *Modification* – binary variable: 1 for yes and 0 for no; indicates whether the program is a modification of a previous program
- *Prototype* – binary variable: 1 for yes and 0 for no; indicates whether the program had a prototyping effort
- *Dem/Val Prototype* – binary variable: 1 for yes and 0 for no; indicates whether the prototyping effort occurred in the PDRR phase

- *EMD Prototype* – binary variable: 1 for yes and 0 for no; indicates whether the prototyping effort occurred in the EMD phase
- *Did it have a PE* – binary variable: 1 for yes and 0 for no; indicates whether the program had a Planning Estimate
- *Significant pre-EMD activity immediately prior to current version* – binary variable: 1 for yes and 0 for no; indicates whether the program had activities in the schedule at least six months prior to MSII decision
- *Did it have a MS I* – binary variable: 1 for yes and 0 for no
- *Terminated* – binary variable: 1 for yes and 0 for no; indicates if the program was terminated

Sipple's initial investigation of the predictor variables reveals that further consolidation of the contractor variables is necessary in order to produce statistically relevant results (Sipple, 2002:65). This stems from the reality that, in the current form, our data lists 45 different individual contractors. This leads to a small number of repeat uses among our programs and produces statistically insignificant results. Sipple overcomes this problem through use of a consolidation matrix, which captures the 1990s cooperate mergers within the industry. See Sipple's (2002) thesis for more information on this topic. Table 4 shows the new category of contractor variables we use for our analysis.

Table 4 - Consolidated Contractors (Sipple, 2002:67)

New List of Contractor Variables
Lockheed-Martin
Northrop Grumman
Boeing
Raytheon
Litton
General Dynamics
No Major Defense Contractor
More than 1 Major Defense Contractor

Sipple develops the maturity variables using the earliest MS II date and the latest MS III date available to compute EMD maturity values that capture the entire EMD phase. This procedure also avoids confusion when multiple MSII and MSII dates are listed for a program. Sipple goes on to describe scarcity problems with certain variables. Specifically, he finds a shortage of usable data points in the EMD maturity variables which use Initial Operational Capability (IOC) or First Unit Equipped (FUE) dates for computation, the *Concurrency Measure Interval*, and the *Concurrency Measure %* (both derived from RAND). Ultimately, the small number of usable data points limits amalgamation of these variables in models.

Preliminary analysis of our data indicates similar scarcity problems. Starting with an initial set of 97 data points, we find that IOC-based maturity variables shrink by 24 data points to 73 usable data points, and more critically, the *FUE – based Maturity of EMD %* and *RAND Concurrency Measure %* reduce to 38 and 39 respectively. Thus, we also recognize the limits of these variables as possible predictors of cost growth due to the shortage of usable data points.

Logistic Regression

As mentioned earlier in this chapter, we build two types of models to accurately predict cost growth. The first model is a logistic regression model. Logistic regression is a special type of regression that predicts a binary or dichotomous response, coded as '0' and '1' (Neter, 1996:567). Figure 3 gives an example of a logistic response function with the dependent variable *R&D (Schedule) Cost Growth* and independent variable *Maturity (Funding Yrs complete)*. From the graph, we interpret the probability of cost growth

decreases as maturity lengthens. We also surmise there is approximately 62.5 percent probability of zero cost growth at a maturity of 10 funding years.

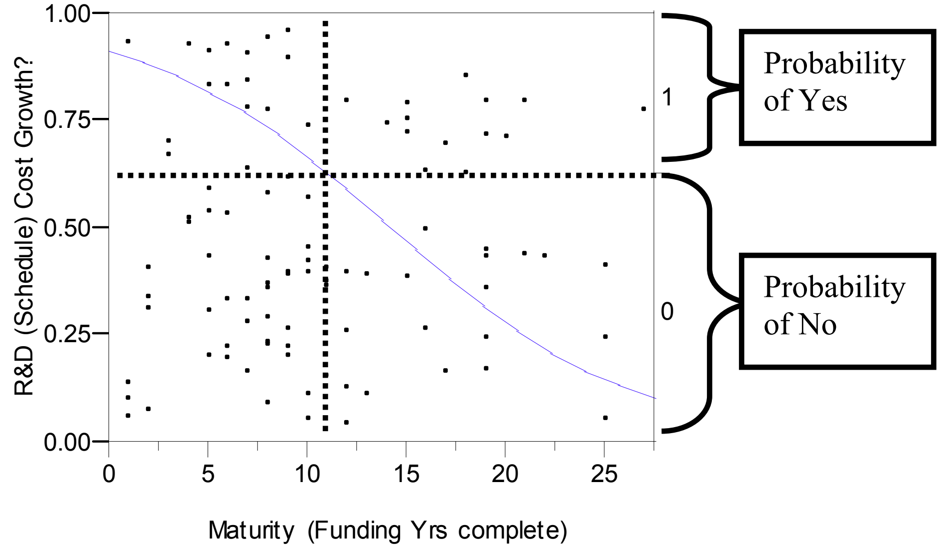


Figure 3 - Logistic Regression Function (JMP Output)

The logistic response function is always constrained by the maximum output values of '0' and '1'. In our case, we search for answers to the question will our program have cost growth or not (yes/no) for each of the SAR cost growth categories under review. In preparation for using logistic regression we add a column to our database which we code a program '1' if it has cost growth (yes) and '0' if it has either zero or negative cost growth (no). Since we now have a distribution of 1's and 0's, we characterize the data as a Bernoulli random variable with probability *p* of success (success=1) (Neter, 1996:568).

The JMP® online help manual further explains the logistic regression process as:

"…the probability of choosing one of the response levels as a smooth function of the factor. The fitted probabilities must be between 0 and 1, and must sum to 1 across the response levels for a given factor value. In a logistic probability plot,

> the y-axis represents probability. For k response levels, *k* - 1 smooth curves partition the total probability (=1) among the response levels. The fitting principle for a logistic regression minimizes the sum of the negative logarithms of the probabilities fitted to the response events that occur–that is, maximum likelihood" (JMP® 5.0, 2002:Help).

Thus, the logistic regression function uses our categorical data to estimate the parameters of a model based upon the "best fit" of the input values. (For more details see Sipple (2002, 68-71)). We use JMP® 4.0 (SAS Institute, 2001) software to accomplish the logistic regression and to build models for estimating whether or not a program will have cost growth.

Since JMP® has no automatic method, equivalent to stepwise regression, for logistic regression, we manually compute thousands of individual regressions, recording our results on spreadsheets. To narrow our search from the approximately 2.6 billion regressions that stem from our 78 predictor variables we observe the following procedure. We investigate all one-variable models for all our candidate variables and record the results. Then we select the nine best models to carry forward for regression using all combinations of two-variable models and record the results. We then select the eight best two-variable models to carry forward for regression using all combinations of three-variable models and record the results. We continue this process, eventually whittling down to the best, most statistically significant, combinations of variables from our pool of predictors. Hence, we call this process the "Darwinist" approach to model development. We stop when we reach a model for which the gain of adding another variable does not warrant the additional complexity of another variable. We find several

candidate models for each number of predictors and then narrow down to the best one for each number of predictors (Sipple, 2002:70-71).

Multiple Regression

The second type model we build to predict cost growth uses multiple regression. As with logistic regression, we use JMP® for the multiple regression analysis. We also utilize the same regression reduction methodology employed during logistic regression to narrow our focus of possible predictor variables. That is we use our Darwinist approach for initial model selection but we also utilize stepwise regression as a backup check to ensure that we have not missed any statistically significant predictor variables from our candidate pool.

Similar to our logistic regression process we find several statistically relevant models exist for each combination of predictors. In each case, we continue model development until we breach our performance measurement of approximately one variable for every ten data points. Using such an approach ensures we do not over-fit the model (Neter, 1996:437).

Ultimately, we seek to construct eight different regression models, which we introduce in this paragraph and expand on in the next chapter. We develop four logistic regression models (one for each SAR cost growth category under analysis) for use with our entire database. These models predict whether a program will have RDT&E cost growth. To simplify our analysis, we call these A models. We then build four multiple regression models (again, one for each SAR cost growth category under analysis) for use with only those programs which experience cost growth. We call these B models, from

which we predict the amount (percent) cost growth the will occur given there is cost growth (from step one). We also apply a log transformation to the response variables in all B model's, in order to correct for heteroskedasticity in the residual plot based on Sipple's (2002) experiences and our own test regressions.

Chapter Summary

This chapter describes the overall research methodology employed during this endeavor. We investigate our source of DoD program information, the SAR database, and describe many of its limitations, as well as some of its benefits. We then discuss our data collection process, and explain our pool of candidate variables. Lastly, we explain the requirement for, and use of, the combination of logistic and multiple regression in our research study. We present the results of our analysis in the next chapter.

IV. Results and Discussion

Chapter Overview

This chapter describes the findings and results of our logistic and multiple regression analysis. We further describe our models and the criteria used to select the final models from the enormous range of possible models. We also analyze the models for statistical validity and applicable use to cost estimators in the field. We intend to conduct analysis using both logistic (A) and multiple regression (B) analysis for each of our four SAR cost growth categories under investigation: Schedule, Estimating, Other and Support. However, as shown later in this chapter, two of the SAR categories – Other and Support have low occurrences of cost growth and do not support meaningful statistical analysis.

Since we eliminate two of the four SAR categories from analysis, our study explores a total of four possible models – one logistic and one multiple regression model for each of the remaining SAR cost growth areas. For identification purposes, we use the first letters of the SAR cost growth category in addition to the alphabetical identification (A / B) of the type of regression model and a numerical number (1- 9) to indicate the generation, or number, of variables associate with a given model. For example, Sch-A3 refers to a Schedule cost growth logistic regression model that has three variables, and Est-B1 refers to an Estimating cost growth multiple regression model that has one predictor variable.

Preliminary Data Analysis

Our research objective seeks to reduce DoD weapons system cost growth through research into the causes of cost growth. With this knowledge, we seek to develop a tool for cost estimators that can effectively estimate cost growth within a program based upon certain program characteristics. As we describe earlier in this thesis, lower risk (uncertainty) equals lower cost growth. We seek to reduce risk via more reliable cost estimates.

The traditional methodology for building a cost growth estimate is with the use of Ordinary Least Squares (OLS) regression techniques. A basic assumption of OLS regression is the underlying data distribution is continuous. However, for our study, the response variable indicates this is not the case. Instead, we find a mixture distribution – a discrete mass at zero and a continuous distribution elsewhere is present. This situation necessitates that we split the data into two separate sets to accurately model the individual effects of both the discrete and continuous data components. As demonstrated by Sipple (2002), a two-step cost growth model produces statistically equivalent results as a single-step regression model however; the two-step model is statistically more reliable due to the validity of its underlying assumptions. For these reasons, we adopt this two-step methodology.

The scope of our research is to fully develop the SAR cost growth categories of Schedule, Estimating, Support and Other. We intentionally, omit the study of the Engineering category since it has been previously studied (Sipple 2002) and the Economic and Quantity categories, by convention, since these are normally excluded during cost growth analysis. We focus only on programs in the SAR that have a DE

baseline during the period 1990 to 2001 and limit our analysis to only the RDT&E appropriation (3600).

Stem and leaf plots of the four cost growth areas under analysis indicate a mixture distribution is present in each category (see Figure 4) and confirms the need for a two-step analytical approach. The mixture distribution is clearly visible by the multiple occurrences of zeros (or no cost growth) centered on zero in the plots. For clarity, we mention that our research treats the negative occurrences of cost growth on these plots as "zero" cost growth in our logistic regression model building and analysis. We observe that the Schedule and Estimating plots appear to have sufficient data to support a meaningful statistical analysis. However, the Other and Support plots appear far less-populated indicating possible small sample populations. We further investigate this possibility, and discover the Other category has only four occurrences of cost growth and the Support category fifteen occurrences (Figure 5). This lack of data points renders these two areas useless for meaningful statistical regression therefore we limit further analysis of these two areas to descriptive measures only.

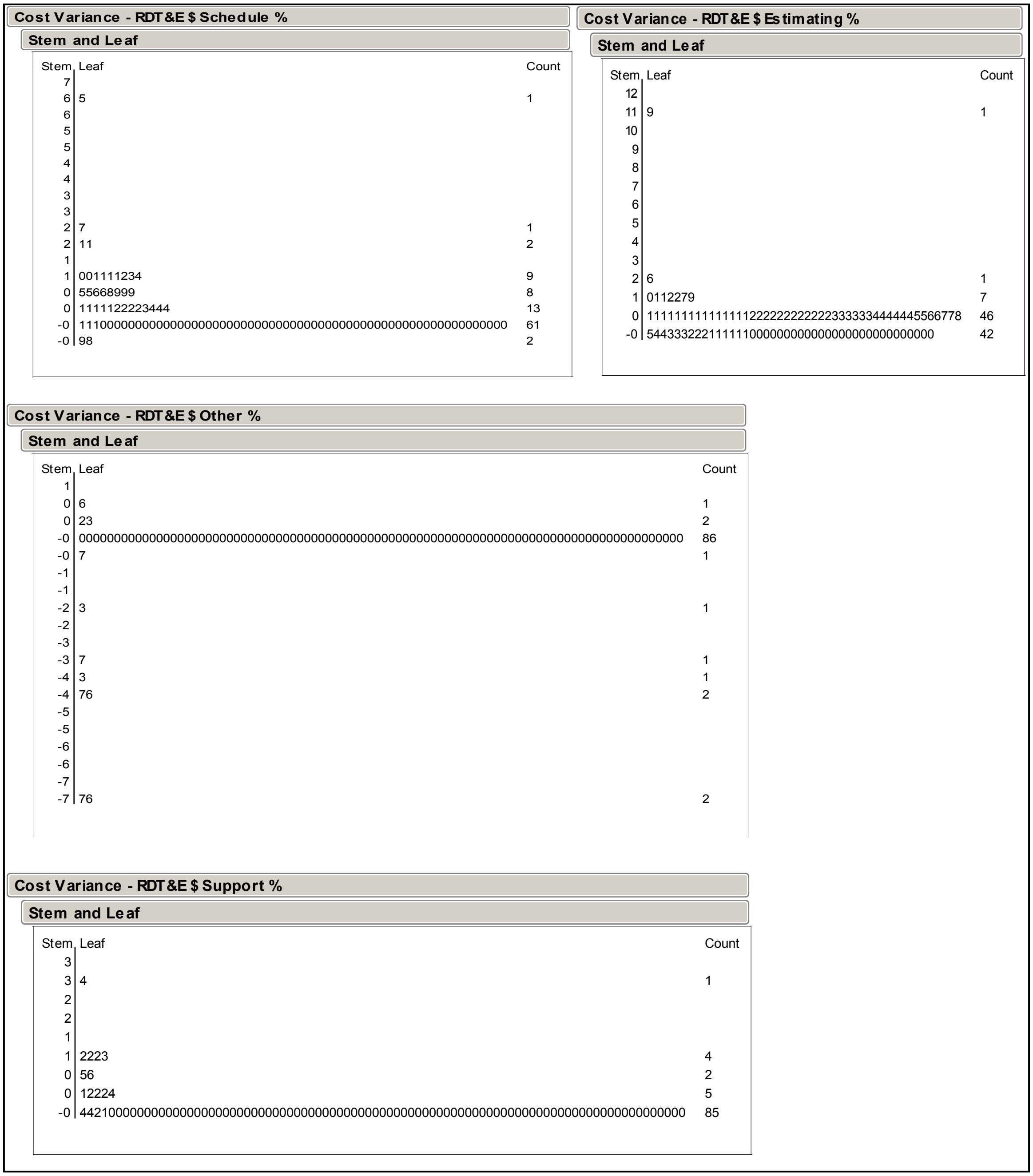


Figure 4 - Stem and Leaf Plots of *Y* Variables (stem in 10's, leaf in 100's)

Analysis of the Other cost growth category reveals the four programs that exhibit cost growth are: B-1B CMUP-Computer, F/A -18 C/D, E-6A TACAMO, and PATRIOT (MIM-104). We find that all four programs exhibit several similar characteristics: all

have “Estimating” cost growth, domain of operation is “Air”, and prototyping or other significant pre-EMD activity occurred in their development.

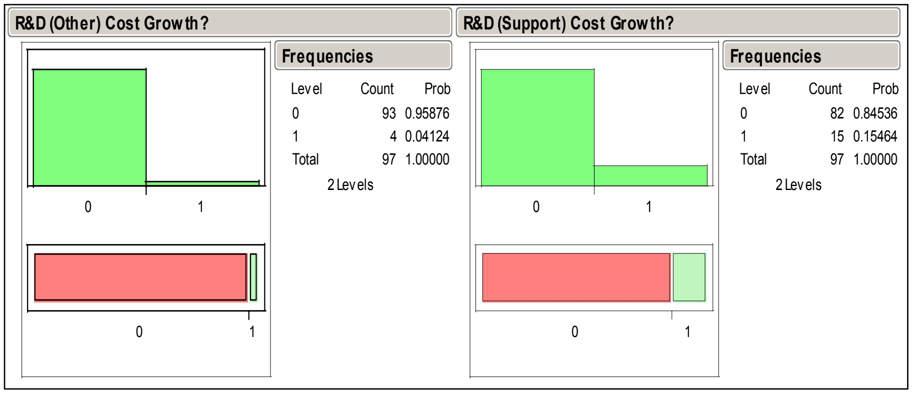


Figure 5 - Frequency Plot of Other and Support Cost Growth

From this information, we gain possible insight into what causes “Other” cost growth but stop short of drawing conclusions based on four programs. We conduct identical analysis for the Support cost growth area but find no commonality between the fifteen programs, which comprise this category. Thus, our analysis of these two SAR cost growth area concludes but we continue on with the Schedule and Estimating categories.

A further visual inspection of the 78 candidate variable plots reveals the existence of two “outliers” in the *New Concurrency Measure %* variable. We must note that our use of the term “outlier” in this instance does not refer to the normal statistical definition of outlier because we are dealing with binary responses and hence do not have customary residual diagnostics to describe the residuals. Thus, we use the term to simply describe data points that can unduly influence the relevance of a variable for model inclusion.

As Figure 6 shows, these data points are significantly separated from the majority of data points. We investigate the effect of these data points on some test models and notice that the test model's R^2 (U) changes from 0.3703 to 0.4808 and the *p*-values of the individual parameters in the model significantly change when the data points are excluded (Figure 7). Since we witness such a fluctuation from the removal of these points, we determine these data points are "influential outliers" and we exclude them from all further logistic regression analysis. Hence, we continue our analysis and model building efforts using the only the Schedule and Estimating categories, and exclude two data points from further model A development. We begin our analysis with the logistic regression models (A) and then move to multiple regression models (B).

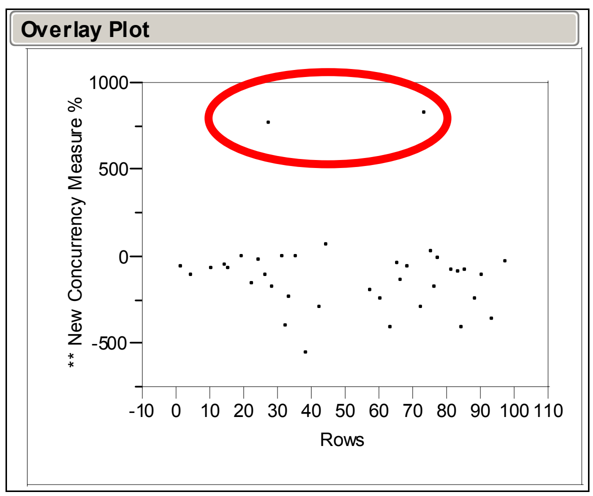


Figure 6 - Overlay Plot of New Concurrency Measure

Nominal Logistic Fit for R&D (Schedule) Cost Growth?

RSquare (U)	0.3703
Observations (or Sum Wgts)	37

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	2.72773267	1.1461739	5.66	0.0173
Maturity (Funding Yrs complete)	-0.2376123	0.0890709	7.12	0.0076
Electronic	2.37165577	1.1235268	4.46	0.0348
Service = AF only	-3.1437756	1.4228221	4.88	0.0271
** New Concurrency Measure %	-0.0041864	0.0018255	5.26	0.0218

Nominal Logistic Fit for R&D (Schedule) Cost Growth?

RSquare (U)	0.4808
Observations (or Sum Wgts)	35

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	1.69362835	1.3565459	1.56	0.2119
Maturity (Funding Yrs complete)	-0.2307171	0.1005855	5.26	0.0218
Electronic	3.61862193	1.5060579	5.77	0.0163
Service = AF only	-3.7930794	1.6912753	5.03	0.0249
** New Concurrency Measure %	-0.0098235	0.0045787	4.60	0.0319

Figure 7 - Logistic Regression Models With and Without Influential Data Points

Logistic Regression Results – Model A

As we discuss in chapter III, we face a staggering manual task of finding the "best" cost growth model from an estimated 2.6 billion possible combinations of models, which originate from our 78 candidate predictor variables. Until recently, our statistical software package, JMP®4.0, offered no automated stepwise-type function for logistic regression to help reduce this task, so we pursued a manual Darwinist approach in selecting our candidate variable models. This methodology selects only the strongest, most statistically significant, models to be carried forward for each successive generation of model building, and culminates with only those combinations of variables (models) surviving which have the most value in predicting cost growth.

However, we discover the newly released JMP® 5.0 offers the additional capability of step-wise for logistic regression. Since we learn of this feature after our initial process has begun, we decide to test this feature to help us quickly obtain a significant predictive cost growth model. We start by adding all 78 variables to the automated step-wise function and immediately find that we exceed the software's capacity for the number of variables used at one time. Next, we try multiple batches of smaller groups so that the automated model runs properly and adjust the sensitivity of the

stepwise model to mirror our manual criteria. We record the ten "best" single variable models that step-wise identifies with the first generation of manual models we previously computed. We find that stepwise does not compare favorably with our manual process. Out of our ten "best" first generation manual models, stepwise identifies only four of those same models. Thus, stepwise fails to identify six of the most significant variables from our candidate pool. We then test if stepwise will identify our "best" four variable manual model also previously computed. We input the four variables along with 20 additional variables into the stepwise model. We find that stepwise does not identify the same four variables nor does it match the level of significance (R^2 (U)) we obtain in our manual model. Furthermore, the stepwise identified four variable model has a lower R^2 (U) than our manual generated four variable model.

From this, we conclude that stepwise can save us significant computational time in reducing the number of variables we consider; however, the trade-off is that our final model will not be as significant as our manually generated model. Thus, we choose to proceed with our initial manual process of model development. We follow this strategy for both the Schedule and the Estimating cost growth models. We commence with a single-variable model and progress to a nine-variable model for the Estimating model.

We further elaborate on the Darwinist approach of our manual model building to give potential end-users an understanding of the magnitude and meticulous detail given to this process. We begin by computing all one-variable models and recording the results on spreadsheets. We select the best nine, one-variable models to carry forward. We regress each of the nine best one-variable models against all 78-candidate predictor variables and record the results. We then select the eight best two-variable models from

these results, and carry these forward for regression using all possible combinations of three-variable models. We continue this process until the advantage of adding variables is outweighed by the additional complexity of another variable. We repeat this process for both our Schedule and Estimating cost growth models, which culminates in approximately nine thousand regressions. For each category of model, and at each generation, we scrutinize and compare several possible candidate models before selecting the best model. Our final selection is based on the optimal mix of statistical measures listed in Table 5. A discussion of these measures follows.

Table 5 - Evaluation Measures for Model A

Measure
R^2 (U)
Number of Data Points / Ratio
Area Under ROC

Our first statistical measure for comparison of models is R^2 (U). The logistic regression R^2 (U) is not the same as the R^2 for ordinary least squares regression. R^2 (U) values range from zero “0” to “1”, and represents the proportion of the total uncertainty that is attributed to the defined model (JMP® 5.0, 2002: Help). The OLS R^2 refers to the amount of variance explained by the regression line, while the logistic regression R^2 (U) is the proportion of variance explained by a dichotomous or categorical dependent variable (Garson, 2003:9). Mathematically, our software, JMP® 5.0 calculates the R^2 (U) statistic as the difference of the negative log likelihood of the fitted model minus the negative log likelihood of the reduced model, divided by the negative log likelihood of the reduced model or simply (JMP® 5.0, 2002: Help):

$$\frac{\text{- loglikelihood for Difference}}{\text{- loglikelihood for Reduced}}$$

Thus, we consider R^2 (U) as a measure of the amount of certainty explained by our model, and recognize that a higher R^2 (U) indicates a better prediction model. See Sipple (2002) for more information on this performance measure.

The second measure we consider in evaluating models is the number of data points. The number of data points available is critically important because the higher the number of data points, the more representative our sample is of our underlying population. Thus, we favor models with the highest number of observations possible when making our model selections. A further benefit of large observations is the ability to add more predictor variables to our models before the model becomes unstable. A basic rule of thumb when selecting the number of variables for model inclusion is that a model should have at least six to ten data points for every predictor variable (Neter, 1996:437). For our research, we immediately exclude any model, which falls below the 6:1 ratio, and cautiously evaluate those models with a ratio between 6:1 and 10:1.

Next, we consider the area under the Receiver Operating Characteristic (ROC) curve as a discriminator between models. The ROC curve is a graphical representation of the relationship between true-positives and false-positives. The curve is a plot of sensitivity by (1 – sensitivity) for each value of *X* where, sensitivity is the probability that *X* correctly predicts the existing condition (true positive) and (1 – sensitivity) is the probability that *X* correctly predicts a condition that does not exist (false positive). If a test was 100 percent accurate (true positive - sensitive), it would pass through the point (0,1) on the ROC grid (see Figure 8) (JMP® 5.0, 2002:Help). Thus, the closer the ROC

curve comes to this point, the higher its ability to predict. Moreover, the larger the area under the ROC curve, the more accurate a model it is at predicting.

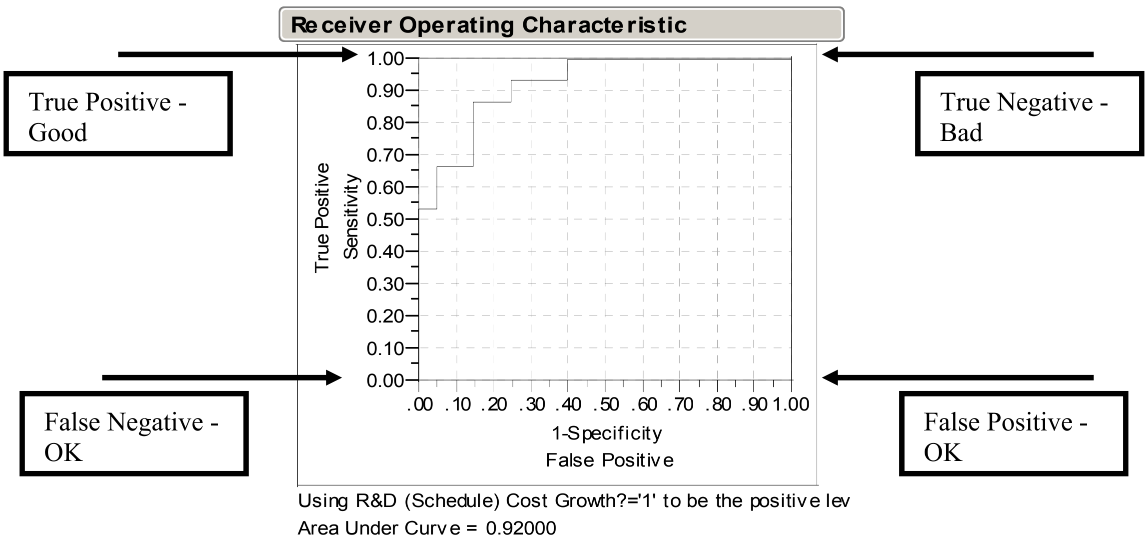


Figure 8 - Receiver Operator Characteristic Curve

For our research, we interpret the ROC curve as the probability of correctly obtaining a true positive when the underlying question is true. In our study, the underlying question is "does my program have cost growth?" A true positive is obtained when a model correctly predicts cost growth in a program that actually has cost growth, and a false positive is obtained when the model predicts cost growth when there is none. We note that a false positive is not a "bad" prediction when referring to cost growth, although a true negative would be "bad" in terms of cost growth estimation. Thus, when evaluating models by this criterion, we search for the model with the largest area under the ROC curve for each category of model, and within each generation of model, we evaluate. See Sipple (2002) for more information on this performance measure.

Table 6 and Table 7 show the results of the Schedule cost growth A model development. Our analysis uncovers two predominate families of models within this category, one which maintains a very high number of data points (95) and a second, which is considerably less overt (35 data points). Yet, the smaller family model has a significantly higher R^2 (U) and a larger area under the ROC curve than the larger family model, indicating more accuracy. We discover the second family after three initial generations of models hence, the reason behind the empty cells in the second model.

Table 6 - Schedule Model A - Performance Measures

Schedule Cost Growth Logistic Regression Models (N=95)

#1	Number of Variables					
	1	2	3	4	5	6
RSq (U)	0.1512	0.2016	0.2547	0.2835	0.3285	0.3463
# Observations	95	95	95	95	95	95
Area Under ROC	0.72452	0.78548	0.82429	0.83405	0.8569	0.8681
Incremental increase of R^2 (U)	0.1512	0.0504	0.0531	0.0288	0.0450	0.0178
Incremental increase under ROC	0.72452	0.06096	0.03881	0.00976	0.02285	0.0112
Ratio: # Obs to variables	95.0	47.5	31.7	23.8	19	15.8

Schedule Cost Growth Logistic Regression Models (N=35)

#2	Number of Variables					
	1	2	3	4	5	6
RSq (U)				0.4808	0.4809	0.5982
# Observations				35	35	35
Area Under ROC				0.92000	0.92000	0.94333
Incremental increase of R^2 (U)				0.4808	0.0001	0.1173
Incremental increase under ROC				0.92000	0.00000	0.02333
Ratio: # Obs to variables				* 8.75	* 7.0	** 5.83

* Caution Zone
** Critical Zone

We recognize that the smaller quantity family (35) immediately breaches the cautionary zone for our ratio of data points to variables, yet we continue with our analysis for two more generations. We progress from the fourth to the fifth generation of models because none of the performance criteria, for the large family (95) models, suggests we have exhausted the benefits of adding extra variables to the models. However, we

observe that one of the large family variables (*RAND Prototype*) *p*-value exceeds 0.05 (Table 7).

Just as in OLS regression, the lower the *p*-value of the parameter estimate, the higher the statistical significance of that parameter in predicting the response variable. For our research, we desire a model with all *p*-values less than 0.05 so that our models are as effective as possible in estimating cost growth. Yet, we are unable to consistently meet this desire throughout our Schedule cost growth model building process. Thus, we ease this restriction to accept *p*-values of up to 0.1.

Table 7 - Schedule Model A - Predictors

Schedule Cost Growth Logistic Regression Models (N=95)

#1	Number of Predictors					
	1	2	3	4	5	6
Maturity (funding Yrs Complete)	0.0001	0.0001	0.0000	0.0000	0.0000	0.0000
AR involvement		0.0158	0.0117	0.0112	0.0036	0.0028
Versions Previous to SAR			0.0133	0.0061	0.0036	0.0095
RAND Prototype				* 0.0687	* 0.0773	
Northrup Grumman					0.0246	0.0162
Significant pre-EMD activity						** 0.1072
EMD Prototype						** 0.1039

Schedule Cost Growth Logistic Regression Models (N=35)

#2	Number of Predictors					
	1	2	3	4	5	6
Maturity (funding Yrs Complete)				0.0218	0.0223	0.0359
Electronic				0.0163	0.0166	* 0.0570
New RAND Concurrency Measure%				0.0319	0.0322	** 0.1066
Service = AF only				0.0249	0.0258	
Aircraft					** 0.9793	
Boeing						0.0435
Class S						** 0.1406
N involvement						* 0.0630

* Caution Zone
** Critical Zone

The fifth generation small family is excluded from further consideration due to high *p*-values for the *Aircraft* variable. We then proceed from the fifth to the sixth generation where we encounter multiple occurrences of high *p*-values for our parameter

estimates in both families of models. Thus, we terminate our search after six generations. For validation, we exclude from consideration Sch #1– A6 (95), Sch #2– A5 (35) and Sch #2– A6 (35) because of *p*-value breaches. Since there is such an extreme drop in the number of data points, and a large jump in R^2 (U) between the two families we decide to carry both models forward to validation. We follow this strategy to test the appropriateness of our selection criteria and overall methodology. Hence, we carry forward from this area to validation, Sch #1– A5 and Sch #2– A4, as the most parsimonious and robust models (Appendix A and B).

Table 8 and Table 9 show the results for the Estimating cost growth model A. Our development and analysis of the Estimating cost growth area continues relatively uneventful for nine generations of models. During the sixth through the ninth generation, we encounter several models with high *p*-values, including one instance in which a model's variable exceeds the 0.1 *p*-value criteria. Specifically, we progress from the seventh to the eighth generation in search of a model with the highest measurement characteristics as possible, and because the majority of our performance measurement criteria are positive, we do not stop. The eighth generation moves our ratio of data points to variables into the cautionary zone, and we find that one of our variables – *Fixed Price EMD Contract* exceeds our 0.1 *p*-value restriction. We note the increasing benefit of adding this variable is slight, increasing our R^2 (U) by only 0.0195, and the area under the ROC curve by 0.0098, but we investigate the possibility that an additional variable might reap greater improvements in our model's measurements, so we proceed.

Table 8 - Estimating Model A - Performance Measures

Estimating Cost Growth Logistic Regression Models

#1	Number of Variables								
	1	2	3	4	5	6	7	8	9
RSq (U)	0.1016	0.1680	0.2104	0.2470	0.3235	0.3912	0.4184	0.4379	0.4676
# Observations	95	95	95	95	88	88	88	86	86
Area Under ROC	0.70492	0.75747	0.79725	0.82956	0.86333	0.89389	0.89813	0.90792	0.91818
Incremental increase of R^2 (U)	0.1016	0.0664	0.0424	0.0366	0.0765	0.0677	0.0272	0.0195	0.0297
Incremental increase under ROC	0.70492	0.05255	0.03978	0.03231	0.03377	0.03056	0.00424	0.0098	0.01026
Ratio: # Obs to variables	95.0	47.5	31.7	23.8	17.6	14.7	12.6	* 10.8	* 9.6

* Caution Zone

Table 9 - Estimating Model A – Predictors

Estimating Cost Growth Logistic Regression Models

#1	Number of Predictors								
	1	2	3	4	5	6	7	8	9
Length of R&D in Funding Yrs	0.0020	0.0007	0.0005	0.0005	0.0002	0.0001	0.0001	0.0001	0.0001
SVS>3		0.0078							
Version Previous to SAR			0.0282	0.0328		0.0094	0.0134	0.0070	0.0044
N involve-ment?			0.0106	0.0026	0.0007	0.0007	0.0009	0.0010	0.0109
PE ?				0.0477	0.0070	0.0100	0.0071	0.0045	0.0031
RAND Lead Svc = DOD					0.0034	0.0038	0.0068	0.0090	0.0060
Did it have a MSI					0.0205	* 0.0914	0.0464	0.0421	* 0.0646
RAND Prototype							* 0.0888	0.0491	0.0436
Fixed-Price EMD Contract								** 0.1268	* 0.0832
SVS>2									* 0.0908

* Caution Zone
** Critical Zone

At the ninth generation, we recognize an incremental improvement in R^2 (U) of 0.0297 and area under the ROC curve of 0.0102, both of which are higher than the contribution of the eight variable but is not the breakthrough we had hoped for. We recognize at this juncture, that with nine variables our model is fairly complex, and that we have multiple variables with less than significant contributions to the model. Thus, we deem the eighth and ninth generation models to be unacceptable since the benefit of adding the extra variables in terms of R^2 (U) and area under the ROC curve is outweighed by the additional complexity from extra variables. The multiple breaches to the significant *p*-value level of 0.05 further solidifies this decision. Hence, we submit to

validation the Est – A7 model as our most robust model of the Estimating cost growth class (Appendix C).

For validation, we utilize the previously selected 25 random data points set aside prior to model building. The 25 data points constitute 20 percent of the original 122-point data set. The logistic regression validation process consists of regressing each specific model to be validated against the entire 122-point data set. We then save the functionally predicted values ('0' or '1') for each of the validation (25) data points and compare to the actual values. JMP® computes the predicted values by assessing the probability of having cost growth based upon the factors in the specific model. We use JMP®'s default settings, in which a '1' is assigned to any point with a probability of 0.5 or greater and a '0' otherwise. However, we note that these settings can be adjusted to allow the cost estimator greater flexibility in assessing cost growth.

In the Schedule cost growth area, we use all 25 data points in validating model #1 but are not as fortunate with model #2, where we lose 18 data points to missing values (or the absence of predictor variable characteristics in the validation set). The culprit in this model is the *New RAND Concurrency Measure %* variable, which accounts for the loss of all 18 data points. We are not surprised by this fact given that our preliminary analysis indicated that this variable had a shortage of usable data points. The abundance of validation data points for model #1 substantiates our modeling criteria of maintaining the largest number of data points as possible – to better represent the underlying characteristics of the population. Hence, model #1's variables (characteristics) are present in all 25-validation points while model #2's are present in only seven.

Upon validation, we find that model #1 accurately predicts 14 out of the 25 data points for a 56 percent success rate. For model #2, we discover that it accurately predicts six out of seven data points for a success rate of 85.7 percent. Since, the success rate of model #1 is only slightly better than flipping a coin for a 50/50 chance, we recognize model #2 and its enviable success rate as our best model for this category. We surmise that although model #1's characteristics are present in every validation point and model #2's characteristics are less represented in the population, the improved accuracy of model # 2 stems from the higher performance measure statistics. This confirms our model development criteria. Thus, we submit Sch #2 –A4 as our best model for this category. See Table 10 for a summary of all model A validation results and Appendix G for the complete validation analysis.

In validating the Estimating cost growth area, we use 23 of the 25 data points (2 data points are lost due to missing values). We find that Est-A7, accurately predicts 18 of the 23 data points for a success rate of 78.2 percent. We are pleased with these results since the model's characteristics are both well represented in the validation population, and have good predictive capability, as evidenced by the reasonably high success rate. Thus, we are satisfied that Est – A7 is our best model in this category. See Appendix A – C for whole model characteristics for all A models.

Table 10 - Model A Validation Results

Model	# Predicted Correct (Total)	% Accurate (Total)
Sch#1 - A5	14	56.00%
Sch#2 - A4	6	85.71%
Est#1 - A7	18	78.26%

Multiple Regression Results – Model B

We continue our two-step methodology by constructing a model to estimate the amount of cost growth a program will incur when a decision maker knows that a program will have cost growth. We start by returning to our randomly selected pool of 97 data points. In each category of cost growth we study (Schedule and Estimating), we exclude programs that have zero or negative cost growth. For the Schedule cost growth area this leaves us with 36 data points and for the Estimating category 63 data points. Under the two-step methodology, using only the data points which contain positive cost growth should give the model more predictive capability, since there is less "noise" to distort and skew the results.

In this section of analysis, we use the same pool of 78 predictor variables as in the logistic regression analysis however, our *Y* response variables change to *Schedule %* and *Estimating %* since we now seek to predict the amount of cost growth in a program. Each respective *Y* response variable is calculated as a percent increase of cost growth from the DE baseline estimate. We begin by analyzing the Schedule cost growth area and then move to the Estimating cost growth area.

An initial plot of the Schedule data indicates the *Y* response variable does not have a normal distribution. We expect this fact since earlier work in this area by Sipple (2002) found the use of a natural log transformation helpful in accounting for distribution shape and to correct for heteroscedasticity in the residual plots. We confirm the appropriateness of a natural log transformation on the Schedule *Y* response variable using JMP® (Figure 9).

However, we do not find as strong a log normal trend in the Estimating Y response variable as shown in Figure 10 by the low KSL (Kolmogorov-Smirnov-Lilliefors) goodness of fit test result of 0.01 for the log normal fit (JMP®, 2002:Help Index). We investigate the possibility that a log transformed Estimating Y response might apply in this case since we have some knowledge of the benefits of its application in previous cost growth research. Figure 10, shows the log transformed Estimating Y response does not pass the Shapiro-Wilk goodness of fit test at an alpha of 0.05; however, by visual inspection we see the distribution is reasonably normal. Thus, we deem use of the log transformed Y response variable appropriate for use on both the Schedule and the Estimating cost growth areas.

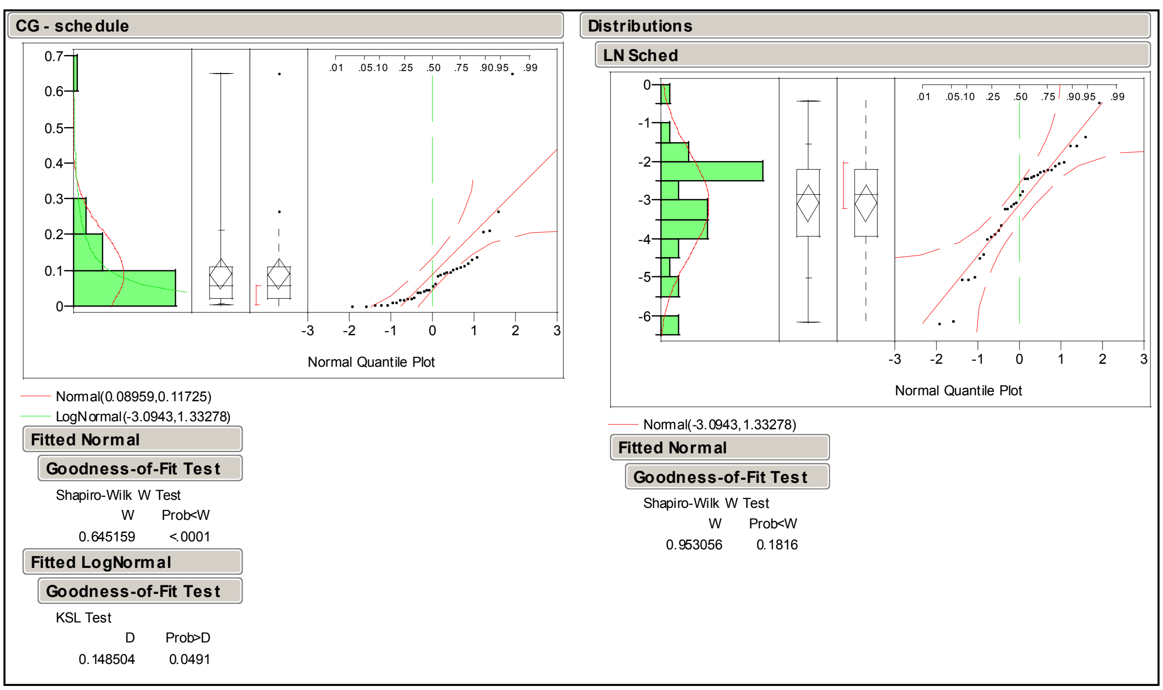


Figure 9 - Distribution of Schedule *Y* and Transformed *Y*

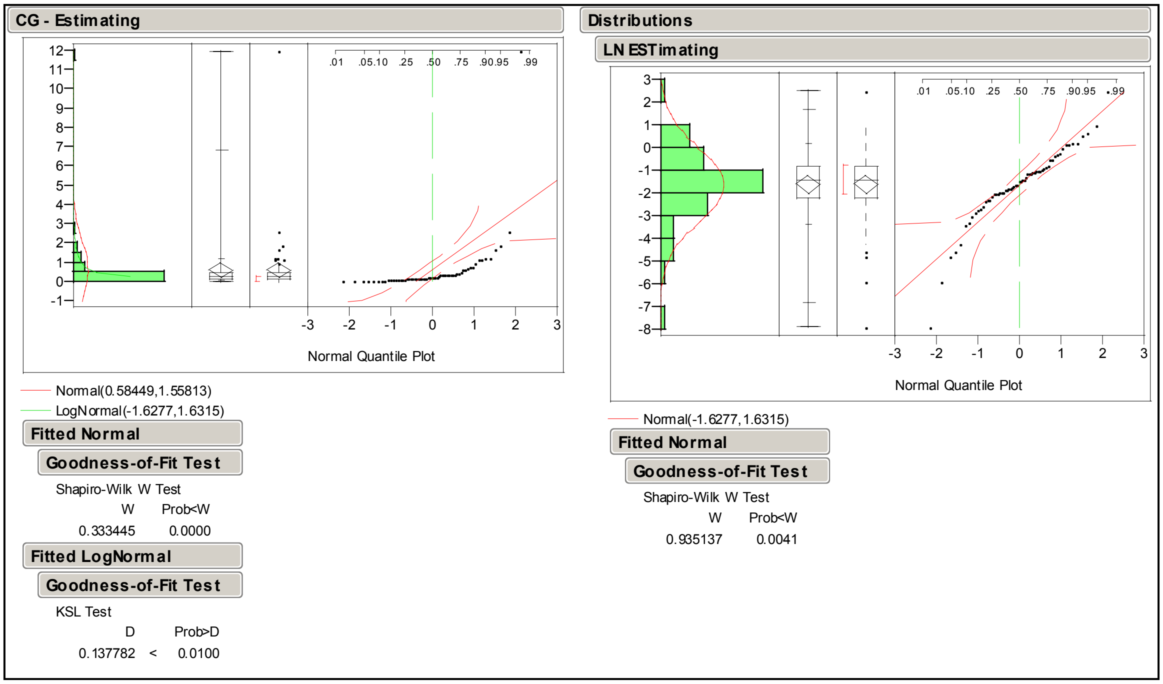


Figure 10 - Distribution of Estimating *Y* and Transformed *Y*

We begin our multiple regression analysis with the Schedule cost growth category. Since this area has only 36 usable data points, we constrain our search for predictive models to only those which contain a maximum of four variables, so that we do not critically exceed our model building benchmark ratio of 10:1 data points to variables. Similarly, we follow the same Darwinist approach to model development that we used during logistic regression.

We initialize the model building process by first, regressing all 78-candidate predictor variables against the Schedule *Y* response variable and record the results on spreadsheets. We then select the top scoring one-variable models and regress against all combinations of two-variable models. We again select the best models and regress against all combinations of three-variable models. We continue this process, searching for the best combination of predictive ability and significant estimates until we breach

one of the model development performance measurements listed in Table 11. The criteria listed in Table 11 are similar to the criteria used for Logistic Regression except that our current focus is on adjusted R^2 instead of R^2 (U). We find the use of adjusted R^2, advantageous over regular R^2, since it protects against artificial inflation of the R^2 value simply by adding additional variables to a model.

Table 11 - Evaluation Measures for Model B

Measure
Adj R^2
Number of Data Points
Ratio: Data Points to Variables

Table 12 and Table 13 display the results of our Schedule cost growth B model development. Our analysis progresses smoothly for two generations of model building. During the third generation, we again discover two predominate models one, which maintains all of its data points (36) – thus, has more prevalent population characteristics, and a second model which has a higher predictive ability, yet contains less prevalent characteristics (27). We are concerned with the smaller model since it immediately reaches a cautionary zone over to its ratio of data points to variables. Because one of its variables is borderline significant at 0.0523 we, however, do not eliminate the model from further evaluation. We proceed to the next generation with two possible models for the Schedule cost growth area. Upon further analysis in the fourth generation, we decide to keep the smaller model despite its aforementioned drawbacks due to its significantly higher adjusted R^2 value compared to model #2. Thus, we carry forward to validation two-candidate Schedule cost growth models (Appendix D and E).

Both of these candidate models pass the statistical assumption tests of normality and constant variance at an alpha = 0.05. We assume independence since there is no obvious serial correlation and we have removed dependent programs from our data set. We further test the predictors for multicollinearity, by ensuring that all variance inflation factors (VIFs) are less than ten (Neter, 1996:387). In fact, all our models VIF's are below 2.0.

Table 12 - Schedule Model B - Performance Measurement

Schedule Cost Growth Multiple Regression Models (N=26)

#1	Number of Variables			
	1	2	3	4
Adj RSq	0.2040	0.4047	0.6081	0.6805
# Observations	36	36	27	27
Incremental increase of R2	0.2040	0.2007	0.2035	0.0723
Ratio: # Obs to variables	36.0	18.0	* 8.7	** 6.8

Schedule Cost Growth Multiple Regression Models (N=35)

#2	Number of Variables			
	1	2	3	4
Adj RSq (U)			0.5597	0.6190
# Observations			36	36
Incremental increase of R2			0.5597	0.0593
Ratio: # Obs to variables			12.0	* 9.0

* Caution Zone

** Critical Zone

Table 13 - Schedule Model B - Predictors

Schedule Cost Growth Multiple Regression Models (N=26)

#1	Number of Predictors 1	2	3	4
Boeing	0.0033	0.0003	0.0002	<.0001
Land Vehicle		0.0012	0.0001	<.0001
RAND Concurrency Measure Interval			* 0.0523	0.0171
Space				0.0208

Schedule Cost Growth Multiple Regression Models (N=35)

#2	Number of Predictors 1	2	3	4
Boeing	0.0033	0.0003	<.0001	<.0001
Land Vehicle		0.0012	<.0001	<.0001
RAND Lead Svs = Navy			0.0012	0.0015
Did it have a MS I ?				0.0204

* Caution Zone

Results from the Estimating cost growth B model development are presented in Table 14 and Table 15. Analysis and model development in this area was by far the most in-depth and extensive out of all the cost growth models and areas we study. From the onset of the second generation, we consider multiple-candidate "best" models and observe the effects on each as we progress through five generations of models. Unfortunately, at the conclusion of our model-building endeavor we disqualified all but one family of models for failure of statistical assumption tests. Even in our surviving best model, we had to remove a data point during the assumptions testing process. The one point we remove was above 0.5 on the Cook's Distance test, indicating it was an influential outlier. In explaining Cook's Distance, Neter has this to say: if the percentile value is less than 10 – 20 percent, the case has little apparent influence on the fitted values, if the percentile value is 50 percent or more, the case has a major influence on the fitted regression (Neter, 1996:381). Thus, to ensure the most reliable, accurate estimates as possible from our models, we are swayed to remove the data point.

Table 14 displays the results of the surviving family of Estimating models. We notice that immediately we loose four data points in generation one, but maintain that level until the third generation. The third generation sees a considerable drop in the number of data points to 45, but an increase of adjusted R^2 to 0.323238. With all the parameter estimates significant, we are encouraged by the possibility of a highly predictive model and progress to the forth generation. We see an increase of .0312 to the Adj R^2 from the addition of two variables: Risk Mitigation and RAND Lead Svc = Navy and the removal one: of General Dynamics. Since all our measurement indicators are positive, and the model parameters continuing to show significance, we proceed to the next generation.

Table 14 - Estimating Model B - Performance Measurements

Estimating Cost Growth Multiple Regression Models

	Number of Variables				
	1	2	3	4	5
Adj RSq	0.1330	0.2482	0.3232	0.3545	0.5225
# Observations	59	59	45	45	44
Incremental increase of R^2	0.1330	0.1152	0.0751	0.0312	0.1680
Ratio: # Obs to variables	59.0	29.5	15.0	11.3	* 8.8

* Caution Zone

Table 15 - Estimating Model B - Predictors

Estimating Cost Growth Multiple Regression Models

	Number of Predictors				
	1	2	3	4	5
Did it have a MS I?	0.0026	<.0001			
Funding Yrs of R&D Completed		0.0029			
IOC - Based Maturity of EMD %			0.0023	0.0013	<.0001
Proc Funding Yr Maturity %			0.0091	0.0096	<.0001
General Dynamics			0.0037		0.0016
Risk Mitigation				0.0014	
RAND Lead Svc = Navy				* 0.0530	0.0033
PE ?					0.0039

* Caution Zone

In this generation we initially see an increase in adjusted R^2 to 0.432103, maintain our 45 data points and notice all *p*-values are highly significant (all less than 0.01). However, at this level, with five variables we reach the cautionary zone of data points to variables, thus we cease further analysis. As mentioned earlier in this section, when we check the statistical assumptions of this model we discover an extreme outlier, which we are obligated to remove. Hence, reducing the total usable data points down by one. When we remove the point, the adjusted R^2 increases to 0.522499 and the number of data points is 44 (as shown in Table 13). Thus, we carry forward to validation the Est – B5 model as the most robust model of the category (Appendix F).

For multiple regression validation, we use the same 25-point validation data set, which we used for logistic regression validation. The validation consists of combining the validation data set with our working data set, and saving the predicted values for each individual model to be validated. JMP® computes the predicted value by fitting the specified model parameters with the values of the 25-point validation set. We then calculate an 80 percent upper prediction bound, back-transform the log normal *Y* response to normal, and assess the accuracy of the model's prediction capability. We utilize an 80 percent upper prediction bound (PB) instead of the traditional 95 percent prediction interval based on Sipple's (2002) work, in which, he finds that after back-transforming the *Y* via the natural exponential function, 95 percent prediction intervals are impractically wide in some cases (Sipple, 2002:87). Hence, the 80 percent attempts to narrow the scope of analysis and ultimately prove more useful to an end user. We gauge the accuracy by comparing the actual percentage cost growth (*Y* response un-

transformed) to the upper prediction bound. A success is recorded when the prediction bound contains the actual value.

For the Schedule category, 11 of the 25 validation data points have cost growth and the other 14 do not. In the Estimating category, 15 of the 25 have cost growth and 10 do not. The percentage of cost growth present in each type cost growth is as follows: Schedule 44 percent and Estimating 60 percent. These distributions seem rationale and representative in light of our working sample population where Schedule cost growth is 36.8 percent (25/95) and Estimating cost growth is 64.2 percent (61/95). Thus, we are unconcerned that every data point is not utilized in the validation step because it well represents the population. In fact, if every data point was used (all contained cost growth) we would be more concerned since this situation would be abnormal.

We begin by validating the Sch#1-B4 (N=27) model with the validation set. We produce our estimates and 80 percent upper prediction bound, and notice that out of the 11 possible programs, 5 have missing values, reducing our usable set to 6 data points. Of these, we produce a prediction bound that accurately captures the true value 66.67 percent of the time with two points falling outside the prediction bound. This result is encouraging (greater that 50/50 chance) however, the small number of observations used to construct the model leaves us a bit uncertain about the widespread application of the model. Table 16 lists the validation results.

For Sch#2-B4 (N=36), we save the predicted values, calculate the prediction bound and find only one missing value (leaving 10 usable). We evaluate and determine an 80 percent success rate with this model, and two data points outside the prediction range. Such results are highly encouraging given the broader base from which this model

originates. This model also aligns well with the statistical premise behind an 80 percent bound, i.e., we expect to see about 80 percent of the validation data points fall below this bound. Thus, we find that this model best serves our purpose of predicting how much cost growth will occur for the Schedule cost growth category. Table 16 shows all the model B validation results.

Table 16 - Model B Validation Results

Model	% of Obs within UB	Usable Pts in Validation	# with Cost Growth	# Missing	# obs used to build
Sch#1 - B4	66.67%	6	11	5	27
Sch#2 - B4	80.00%	10	11	1	36
Est - B5	100.00%	13	15	2	44

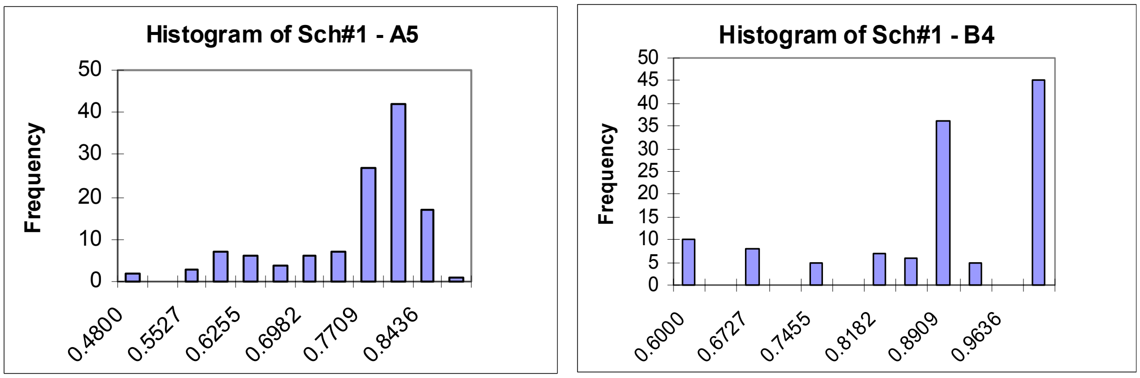
Finally for Est-B5, we calculate our stated values and notice two missing values in the data set, leaving us with 13 usable data points to compute its accuracy. Upon inspection of the 13 estimates, we find a remarkable 100 percent accuracy rate of the actual value being contained by the prediction bound. Since this model was constructed with a large percentage of it original data points 68.8 percent, the most number of data points of any of the multiple regression B models, we are most confident in its results. Such results also seem to add credence to our modeling criteria specifically, maintaining the largest possible number of observations and significant parameter *p*-values. See Appendix H for all the model B validation results.

Rolling Validations

Since Sch #1 - A5 validated at only 56 percent accuracy and Sch#1 – B4 validated at 66 percent accuracy, we investigate the use of a rolling validation window; otherwise known as “Jackknifing” to better evaluate these models’ true predictive capability. We

do this by comparing each models actual cost growth data to either the logistic regression predicted value (1/0) or the back-transformed 80 percent upper bound for all 122 data points versus just for the 25 point validation set. First, we take data points 1 – 25 (validation set) and calculate the accuracy rate for this group. Next, we take data points 2-26 (2-24 from the validation set plus 1 data point from the original data set) and compute the accuracy. We continue this successive process until we have rotated through the entire 122 data points. Lastly, we compute the average and standard deviation for the entire process and graph the results for each respective model.

From Figure 11 we see that Sch#1 - A5 achieves an average 74.59 percent accuracy rate when compared over the entire 122-point data set and Sch # - B4 achieves an average 87.30 percent accuracy. Figure 11 also, shows histograms of the grouped accuracy rates for each model under review. The Sch#1-A5 model shows the true distribution is highly skewed left indicating a strong possibility for lower accuracy predictions on average. Sch#1 – B4's plot shows a choppy distribution with large occurrences of high accuracy on the right side of the graph, low frequency in the middle and medium frequency on the right, producing a slight bath tub shape. This shape suggests that on average the model will predict accurately but we can expect some variation in results (see standard deviation).



Avg	**0.7459**	**Avg**	**0.8730**
Std Dev	**0.0926**	**Std Dev**	**0.1284**

Figure 11 - Jackknife Results

Overall, these results indicate that on average these models will perform reasonably well but incremental performance may be sub-par. For example, Sch#1 – A5 has an average accuracy rate of almost 75 percent yet it predicts below the 50 percent accuracy rate on a few occasions. From this we realize, our initial validation scores are due to random chance and the best application for these models is with large data sets. Thus, we keep with our original selections as the “best” models discussed earlier in this chapter.

Chapter Summary

This chapter elaborates on our model development process, and describes the results from our analysis using these models. We further authenticate the motive for using a two-step methodology consisting of: first, logistic regression to predict if a program will have cost growth and then second, multiple regression to determine how much cost growth will occur, based on the composition of our database. We delve into the criteria and selection process used to establish both types of predictive models, and

assess the usefulness and accuracy of the models using a 25-point validation data set. From these results, we evaluate and select the best model from each category studied and present to the reader for scrutiny.

Our analysis shows that to predict "if" a program will have cost growth from within the Schedule category of cost growth, model #2-A4 is preferred and within the Estimating category, model A7 is preferred. To predict the amount of cost growth, we find that model #2-B4 is the most desirable in the Schedule category, and model B5 is preferred when in the Estimating cost growth arena. A final discussion and application of these models is presented in the next chapter.

V. Conclusions

Chapter Overview

This chapter reviews the pressures that exist in the DoD acquisition environment of major weapons systems procurement and which underscore the necessity of this research. We explore previous cost growth research to investigate the causes of cost growth and for edification of historical or traditional methods of calculating cost growth. We discuss the limits, application and benefits of this research to the DoD cost estimating community, and assess our results with our initial research objective of reducing DoD weapons system cost growth. Lastly, we present several possible follow-on topics to this research.

Restatement of the Problem

Two central problems face the DoD acquisition community today – reduced funding and escalating costs. Excluding recent growth due to the War on Terrorism, the DoD budget declined 29.18 percent from 1985 to 2001. This substantial decrease in budget size restricts current DoD acquisition programs and severely limits the growth of new programs. Reduced funding levels exacerbate the second problem of spiraling major weapons system program cost and program overruns. In fact, we find the average DoD major weapons system program experiences 20 plus percent cost growth from the time of start-up to full-scale production (Drezner, 1993:xiii; Coleman, 2000:19-20).

These two opposing forces have a direct and negative impact on the cost estimators' ability to deliver accurate, consistent and reliable program cost estimates. Our research seeks a partial solution to this problem. Obviously, our study cannot

influence the Congressional budget process or stabilize funding of major weapons system programs. But we can develop a tool to improve accuracy and reliability of cost estimates thus, limiting and perhaps preventing acquisition program cost growth. Specifically, our research develops a unique two-step statistical model to predict cost growth. Our model provides the cost estimator with a quantitative tool to estimate program costs early in a programs acquisition lifecycle. Our estimating tool is more reliable, based on quantitative methods, than subjective cost estimating methods normally available early in a programs life cycle. Thus, as reliability increases, uncertainty about the program decreases, and cost growth (or cost risk) is reduced.

Limitations

We set out to predict cost growth for the Schedule, Estimating, Support and Other SAR cost growth categories, but discover insufficient cost growth data to support inferential analysis of the Support and Other cost growth areas. This limits our research to descriptive measures only for the Support and Other cost growth areas yet, does not hamper a complete inferential analysis of the Schedule and Estimating SAR cost growth areas.

We build our models from historical SAR reports of DoD acquisition programs between 1990 and 2001. We include only programs with a DE baseline estimate falling within this time period and focus exclusively on RDT&E funds. Hence, we are further limited by these boundaries in the use and application of our results. Lastly, we caution the reader against extrapolation of our results beyond the aforementioned bounds used to

develop them. Use of these models beyond these confines may produce erroneous results.

Review of Literature

We perform a review of recent literature on cost growth within the DoD. We find many studies that explore the roots causes of cost growth, as well as, seek to predict cost growth with regression models. Many studies use SAR reports as the source data from which they compute cost growth. Consequently, we find many similarities between the (historical) literature review studies and individual elements of our research effort however; we find only one study that parallels ours in scope. Sipple (2002) focuses on cost growth of RDT&E funded programs that use a DE baseline estimate, and predicts the SAR cost growth category of Engineering. In addition, Sipple assembles a pool of 78 predictor variables extracted from twelve historical cost growth studies. The near identical match between Sipple's (2002) research and ours leads us to the conclusion we can effectively pattern our methodology on Sipple's findings. That is, we benchmark Sipple's predictor variables, procedures and overall methodology for use in our research.

Review of Methodology

Our two-step methodology of predicting cost growth is new to the cost estimation field. The two-step methodology, introduced by Sipple (2002), establishes the use of first, logistic regression to predict "if" a program will have cost growth and second, if applicable, multiple regression to estimate the amount of cost growth expected. This process is new because the traditional (historical) method of predicting cost growth originates around a single-step regression process.

We build upon Sipple's (2002) existing SAR database comprised of major acquisition programs from all service components, which use a DE baseline estimate. The database contains both RDT&E and procurement dollar programs that have an EMD phase of development between 1990 and 2000, to which, we add calendar year 2001 programmatic data. This research focuses strictly on RDT&E dollar accounts yet, we collect procurement dollars information in our process to amass a comprehensive database and to allow for possible follow-on research. (See the last section of this chapter for further follow-on topics.) We convert all programmatic dollar amounts into a common base year (2002) and compute our response variables. Since our database contains a mixture distribution, a point mass of data centered on zero and continuous elsewhere, we split the data into two parts (discrete and continuous) and model each independently. This database contains 122 total data points of which 25 data points (20 percent) are set aside for validation, leaving 97 data points (80 percent) for model development.

We first, compute the logistic regression *Y* response variable *R&D Cost Growth?* for each of our SAR cost growth categories (Schedule, Estimating, Support and Other) to model the discrete data. These variables represent the binary response to the question "does my program have cost growth?" where 1 equals "yes" and 0 equals "no." Next, we compute the multiple regression *Y* response variables - *Schedule %, Estimating%, Support %, Other %* for use with the continuous data. These variables represent the total cost variance (in RDT&E dollars) divided by the respective DE baseline estimate, and answer the question "how much cost growth will occur?" For identification purposes, we call the logistic regression model (A) and the multiple regression model (B).

We investigate the response variables and discover the Support and Other category do not have sufficient data to support inferential statistical regression. Thus, we limit analysis of these two areas to descriptive measures only. We also discover that we must use a log normal transformation on the model B *Y* response variables to correct for heteroscedasticity in the residual plots. The use of the log transformed *Y* response ensures that the underlying assumptions of OLS regression are met.

Development of models A and B employ the Darwinist variable selection strategy described earlier in this thesis and culminate in a pool of candidate "best" models for each category under investigation. We authenticate the single "best" model from the pool of candidate models with our validation data set. We also perform a further statistical investigation of two models to confirm the true accuracy rate using the "Jackknife" procedure of resampling.

Restatement of Results

Our analysis finds that predicting "if" a program will have cost growth (model A), in the Schedule category of cost growth, model #2-A4 is preferred (Appendix B). This model accurately predicts approximately 85 percent of the validation data and all four predictor variables are significant with *p*-values less than 0.05. In the Estimating cost growth category, model A7 (Appendix C) accurately predicts approximately 78 percent of the validation data. Four of the seven predictor variables are highly significant with *p*-values below 0.01, and two of the remaining three variables are below 0.05.

We find that when predicting the "amount" of cost growth a program will experience (model B) in the Schedule cost growth category that model #2-B4 is the most

desirable. This model accurately predicts 80 percent of the validation data and all four predictor variables are significant with *p*-values less than 0.02. In the Estimating cost growth category, model B5 accurately predicts 100 percent of the validation data and three of the five predictor variables are highly significant with *p*-values <0.0001, and the remaining two variables *p*-values are less than 0.02.

Recommendations

Our research confirms the appropriateness of logistic regression in DoD cost analysis, and substantiates the aptness of the two-step methodology to predict cost growth in DoD major weapons systems acquisitions. Use of logistic regression and the two-step methodology provide cost estimators a tool to accurately estimate the cost of weapons programs while it improves reliability of the cost estimate. Such steps support Congressional and Presidential direction to calculate the true or "realistic cost" of DoD acquisition programs.

Logistic regression predicts a binary or dichotomous response. When used in conjunction with OLS regression (and the *Y* response is log transformed), as in our two-step model, it acts as a filter to remove noise or bias from the data stream. The result is a clear, more reliable picture of a weapons system program cost. Moreover, use of logistic regression allows cost estimators to specify a percentage level of certainty for the predicted outcome. For example, a conservative approach might set the model controls at 25 percent or more = "yes", otherwise "no" (the lower the level is set the more likely the model is to predict cost growth, and the higher the initial estimate, due to the increased prediction, the lower cost growth will be). This flexibility allows cost estimators to

adjust the sensitivity level or conservativeness of each individual estimate as necessary to meet program requirements and used responsibly adds to the presidential call for more realistic estimates.

This research demonstrates the effectiveness of logistic regression and OLS regression to predict DoD weapons system cost growth. Logistic regression which predicts if a program will have cost growth (yes/no) and, when applicable (yes responses), OLS regression predicts the amount of cost growth expected. Clearly, the advantages and benefits of this model warrant its implementation for use across the DoD in estimating major weapons system program costs. We further submit that use of logistic regression has a wider place within the DoD community that is as yet unrecognized. For example, logistic regression is used extensively, and successfully, in other industries like the medical occupation career field to predict (true/false) infectious diseases. The DoD should learn from this civilian industry practice and adopt the use of logistic regression not only for major weapons system cost estimates but also for day-to-day cost analysis decisions.

Possible Follow-on Theses

We recommend further cost growth analysis using the two-step methodology demonstrated by this research, as well as, exploitation of our extensive database. Although, this research completes the study of the individual SAR cost growth categories within the RDT&E area, there are several other possibilities for meaningful research. For example:

- Calculate the overall RDT&E cost growth and compare with the combined results obtained from our thesis and Sipple's (2002).

- Calculate individual SAR category cost growth for the procurement accounts within the EMD phase.

- Calculate a combined cost growth estimate for the RDT&E and procurement accounts within EMD.

- Compare individual RDT&E cost growth with individual procurement cost growth. Identify trends, accuracy and root causes within each category.

- Compare overall RDT&E cost growth with overall procurement cost growth. Identify trends, accuracy and root causes.

- Expand methodology to other phases of acquisition (PDRR and procurement). Develop predictive "forecast" variable to link cost growth between phases.

Chapter Summary

This research, combined with Sipple's (2002), presents a solid picture of the drivers of EMD cost growth and develops associated tools for predicting cost within this arena. We investigate thousands of individual regressions to find the germane characteristics that drive cost growth in the SAR cost growth areas of Schedule and Estimating, and develop two models A and B for predicting cost growth. We show that the two-step methodology is required due to the composition (mixture distribution) of our data, and that such a process produces meaningful, reliable statistical results from which accurate cost estimates can be derived.

Appendix A – Schedule Cost Growth Five Variable A Model

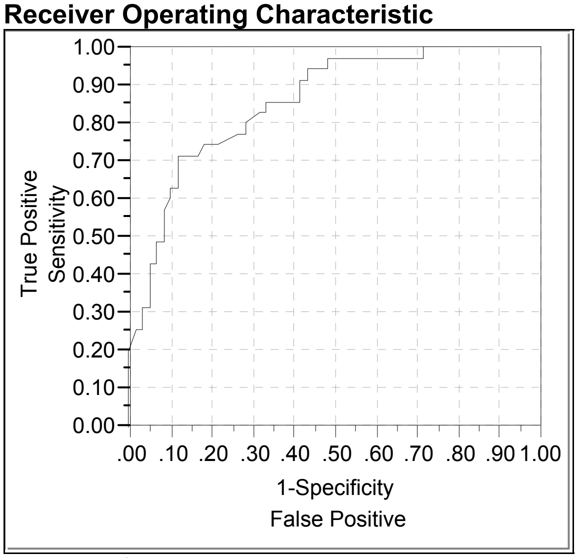
Nominal Logistic Fit for R&D (Schedule) Cost Growth?

RSquare (U)	0.3285
Observations (or Sum Wgts)	95

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	2.97536859	0.7519593	15.66	<.0001
Maturity (Funding Yrs complete)	-0.2364067	0.0552939	18.28	<.0001
AR Involvement?	1.91631441	0.6591188	8.45	0.0036
Versions Previous to SAR	-1.811036	0.6220835	8.48	0.0036
RAND Prototype?	1.05138325	0.5951769	3.12	0.0773
Northrop Grumman	-2.3214554	1.0328553	5.05	0.0246

Receiver Operating Characteristic



Area Under Curve =
0.85690___

Appendix B – Schedule Cost Growth Four Variable A Model

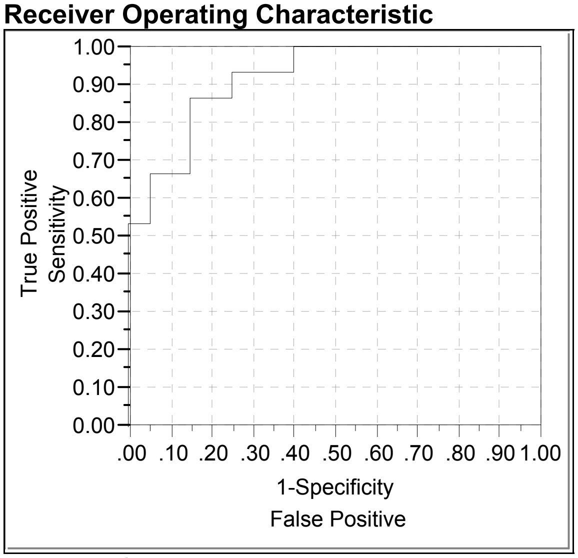
Nominal Logistic Fit for R&D (Schedule) Cost Growth?

RSquare (U)	0.4808
Observations (or Sum Wgts)	35

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	1.69362835	1.3565459	1.56	0.2119
Maturity (Funding Yrs complete)	-0.2307171	0.1005855	5.26	0.0218
Electronic	3.61862193	1.5060579	5.77	0.0163
New RAND Concurrency Measure %	-0.0098235	0.0045787	4.60	0.0319
Service = AF only	-3.7930794	1.6912753	5.03	0.0249

Receiver Operating Characteristic



Area Under Curve =
0.92000__

Appendix C – Estimating Cost Growth Seven Variable A Model

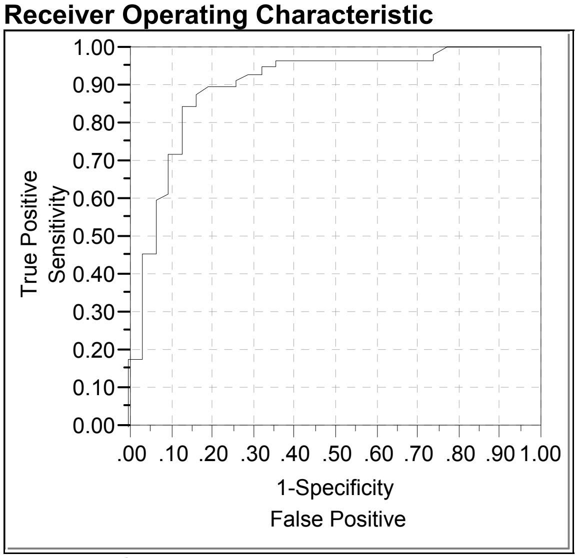
Nominal Logistic Fit for R&D (Estimating) Cost Growth?

RSquare (U)	0.4184
Observations (or Sum Wgts)	88

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	1.73236112	1.0518102	2.71	0.0996
Length of R&D in Funding Yrs	-0.2342976	0.0608878	14.81	0.0001
Versions Previous to SAR	-1.7689849	0.7153578	6.12	0.0134
N Involvement?	2.64212378	0.7978189	10.97	0.0009
Did it have a PE ?	-3.1530811	1.1713143	7.25	0.0071
RAND Lead Svc = DoD	6.49342975	2.4005851	7.32	0.0068
Did it have a MS I ?	1.5486272	0.7774387	3.97	0.0464
RAND Prototype?	-1.1289653	0.6633286	2.90	0.0888

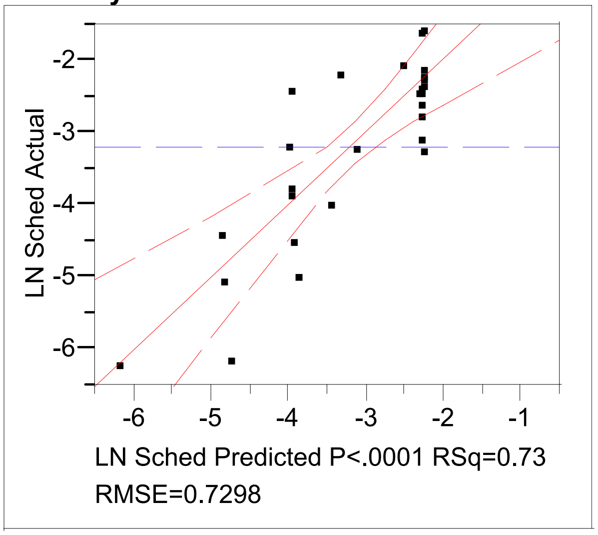
Receiver Operating Characteristic



Area Under Curve =
0.89813__

Appendix D – Schedule Cost Growth #1 Four Variable B Model

Whole Model
Actual by Predicted Plot



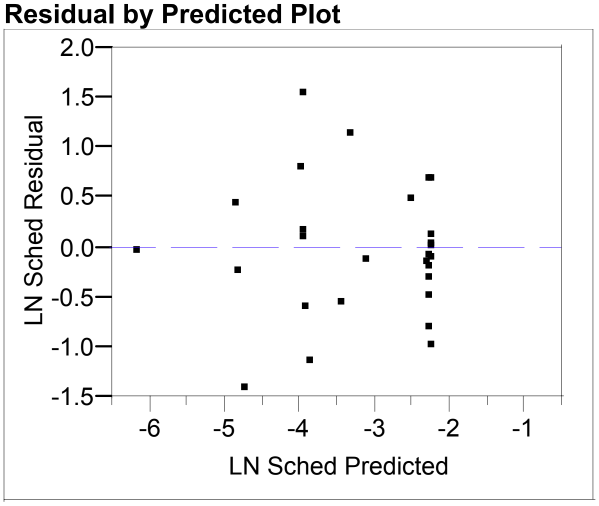
Summary of Fit

RSquare	0.729634
RSquare Adj	0.680476
Root Mean Square Error	0.729794
Mean of Response	-3.20054
Observations (or Sum Wgts)	27

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-2.292496	0.193583	-11.84	<.0001
Boeing	-1.682862	0.312896	-5.38	<.0001
Land Vehicle	-3.925248	0.755936	-5.19	<.0001
RAND Concurrency Measure Interval	0.0008494	0.000329	2.58	0.0171
Space	1.4342709	0.575692	2.49	0.0208

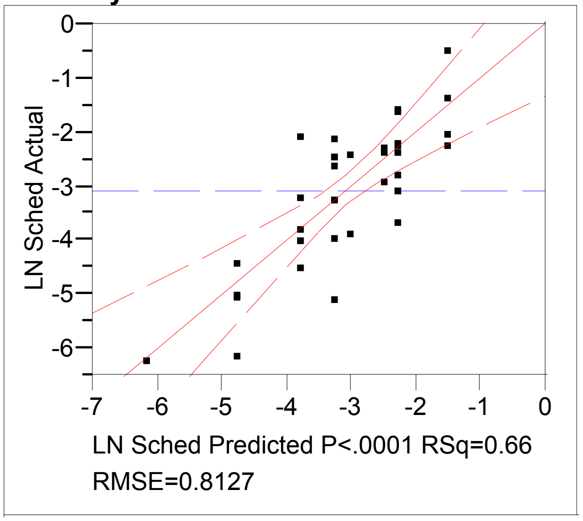
Residual by Predicted Plot



Appendix E – Schedule Cost Growth #2 Four Variable B Model

Whole Model

Actual by Predicted Plot



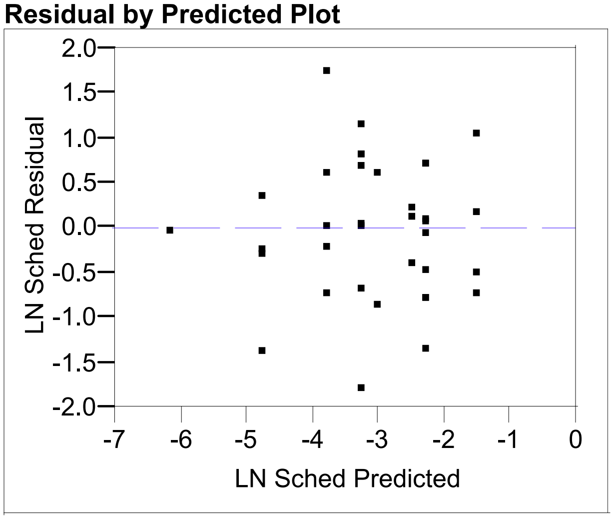
Summary of Fit

RSquare	0.662501
RSquare Adj	0.618953
Root Mean Square Error	0.812668
Mean of Response	-3.07976
Observations (or Sum Wgts)	36

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-1.52265	0.296423	-5.14	<.0001
Boeing	-1.495189	0.299541	-4.99	<.0001
Land Vehicle	-4.660582	0.865041	-5.39	<.0001
RAND Lead Svc = Navy	-0.978341	0.280881	-3.48	0.0015
Did it have a MS I ?	-0.779745	0.318957	-2.44	0.0204

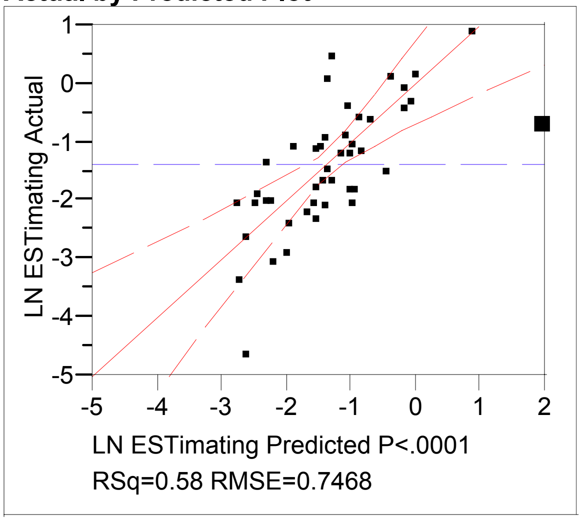
Residual by Predicted Plot



Appendix F – Estimating Cost Growth Five Variable B Model

Whole Model

Actual by Predicted Plot



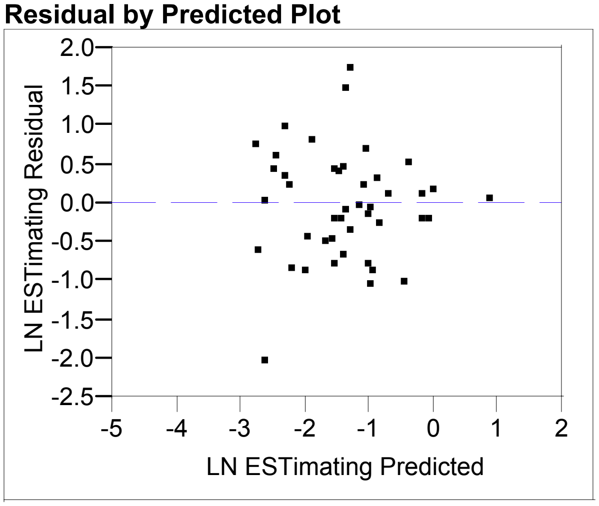
Summary of Fit

RSquare	0.578022
RSquare Adj	0.522499
Root Mean Square Error	0.7468
Mean of Response	-1.3887
Observations (or Sum Wgts)	44

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-1.147983	0.250922	-4.58	<.0001
IOC -Based Maturity of EMD %	0.5759717	0.131413	4.38	<.0001
Proc Funding Yr Maturity %	-1.910945	0.404058	-4.73	<.0001
General Dynamics	-1.282748	0.378116	-3.39	0.0016
RAND Lead Svc = Navy	0.7926428	0.252644	3.14	0.0033
Did it have a PE ?	-0.927261	0.301534	-3.08	0.0039

Residual by Predicted Plot



Appendix G – Model A Validation Results

Sch#1 A5

	Actual LN - Sch	Most Likely ValidSchA#1	Actual Sch Cost Growth	Correct ?
1	-2.096716	1	1	Yes
2	-2.819633	1	1	Yes
3	.	1	0	No
4	-4.065817	1	1	Yes
5	.	1	0	No
6	-1.637677	0	1	No
7	.	0	0	Yes
8	.	0	0	Yes
9	.	1	0	No
10	-1.732971	0	1	No
11	.	1	0	No
12	-1.963355	1	1	Yes
13	.	1	0	No
14	.	0	0	Yes
15	.	0	0	Yes
16	.	1	0	No
17	-2.462601	1	1	Yes
18	-3.330922	0	1	No
19	-1.862817	0	1	No
20	-5.787841	1	1	Yes
21	.	0	0	Yes
22	.	0	0	Yes
23	.	0	0	Yes
24	-7.785513	0	1	No
25	.	0	0	Yes

Counts

14	Yes
11	No
56.00%	Accuracy Rate

Appendix G – Model A Validation Results

Sch#2 A4

	Actual LN - Sch	Most Likely ValidSchA#2	Actual Sch Cost Growth	Correct ?
1	-2.096716		1	N/A
2	-2.819633		1	N/A
3	.		0	N/A
4	-4.065817	1	1	Yes
5	.	1	0	No
6	-1.637677		1	N/A
7	.		0	N/A
8	.	0	0	Yes
9	.		0	N/A
10	-1.732971	1	1	Yes
11	.		0	N/A
12	-1.963355		1	N/A
13	.		0	N/A
14	.		0	N/A
15	.		0	N/A
16	.	0	0	Yes
17	-2.462601		1	N/A
18	-3.330922		1	N/A
19	-1.862817		1	N/A
20	-5.787841		1	N/A
21	.	0	0	Yes
22	.		0	N/A
23	.		0	N/A
24	-7.785513		1	N/A
25	.	0	0	Yes

Counts	
6	Yes
1	No
85.71%	Accuracy Rate

Appendix G – Model A Validation Results

Est - A7

	Actual LN - Est	Most Likely ValidaEstA	Actual Est Cost Growth	Correct ?
1	-0.980085	1	1	Yes
2	.	0	0	Yes
3	.	0	0	Yes
4	-2.137925	1	1	Yes
5	-2.068357	1	1	Yes
6	-3.900665		1	N/A
7	-5.946093	1	1	Yes
8	-3.849291	1	1	Yes
9	.	1	0	No
10	-2.063063	1	1	Yes
11	-2.180035	1	1	Yes
12	-2.054327	1	1	Yes
13	-1.756433	1	1	Yes
14	.	1	0	No
15	.	0	0	Yes
16	-2.877427	1	1	Yes
17	.	0	0	Yes
18	-0.758505	0	1	No
19	-2.266171	1	1	Yes
20	.	1	0	No
21	.	1	0	No
22	-3.212324	1	1	Yes
23	.		0	N/A
24	-0.930631	1	1	Yes
25	.	0	0	Yes

Counts

18	Yes
5	No
78.26%	Accuracy Rate

Appendix H – Model B Validation Results

Schedule#1-B4 Model

Validation	Actual CG - Sch	Predicted LN - Sch	StdErr Indiv LN - Sch	Sch #1 B4 80% UB	Back-T Sch#1B4 80%	Actual within PB?
1	0.122859	-0.849618	0.953497	0.409951	1.506745	Yes
2	0.059628	.	.	.	.	No Data
3		-3.2215	0.817192	-2.14199	0.117421	No Data
4	0.017149	-2.352829	0.754019	-1.35677	0.257491	Yes
5		-3.958087	0.776786	-2.931953	0.053293	No Data
6	0.194431	.	.	.	.	No Data
7		.	.	.	.	No Data
8		-3.95469	0.77697	-2.928312	0.053487	No Data
9		-3.088801	0.798727	-2.033683	0.130853	No Data
10	0.176758	-2.297649	0.754917	-1.300403	0.272422	Yes
11		-2.293374	0.755012	-1.296003	0.273623	No Data
12	0.140387	-7.045929	1.084908	-5.612766	0.003651	No
13		.	.	.	.	No Data
14		-2.308011	0.754702	-1.311049	0.269537	No Data
15		.	.	.	.	No Data
16		-3.808175	0.786957	-2.768604	0.06275	No Data
17	0.085213	.	.	.	.	No Data
18	0.03576	.	.	.	.	No Data
19	0.155235	-4.825031	0.802419	-3.765036	0.023167	No
20	0.003065	.	.	.	.	No Data
21		-2.288164	0.755132	-1.290635	0.275096	No Data
22		-2.261466	0.755834	-1.263009	0.282802	No Data
23		-2.285616	0.755193	-1.288006	0.27582	No Data
24	0.000416	-2.263165	0.755785	-1.264773	0.282303	Yes
25		-3.952085	0.777113	-2.92552	0.053637	No Data

Counts

4	Yes
2	No
66.67%	Obs Within PB

Appendix H – Model B Validation Results

Schedule#2 B4

Validation	CG - Sch (actual)	Predicted LN - Sch 2	StdErr Indiv LN - Sch 2	Sch#2 B4 80% UB	Back-T Sch#2 B4 80%	Actual within PB?
1	0.122859	-1.52265	0.865041	-0.390311	0.676846	Yes
2	0.059628	-3.797584	0.857755	-2.674783	0.068922	Yes
3		-3.280736	0.846293	-2.172938	0.113843	No Data
4	0.017149	-3.280736	0.846293	-2.172938	0.113843	Yes
5		-3.797584	0.857755	-2.674783	0.068922	No Data
6	0.194431	.	.	.	.	No Data
7		-2.302395	0.846266	-1.194632	0.302815	No Data
8		-3.99618	0.912236	-2.802063	0.060685	No Data
9		-2.302395	0.846266	-1.194632	0.302815	No Data
10	0.176758	-2.302395	0.846266	-1.194632	0.302815	Yes
11		-1.52265	0.865041	-0.390311	0.676846	No Data
12	0.140387	-6.962977	1.192724	-5.401702	0.004509	No
13		-1.52265	0.865041	-0.390311	0.676846	No Data
14		-2.302395	0.846266	-1.194632	0.302815	No Data
15		-3.280736	0.846293	-2.172938	0.113843	No Data
16		-3.797584	0.857755	-2.674783	0.068922	No Data
17	0.085213	-3.280736	0.846293	-2.172938	0.113843	Yes
18	0.03576	-2.302395	0.846266	-1.194632	0.302815	Yes
19	0.155235	-3.797584	0.857755	-2.674783	0.068922	No
20	0.003065	-3.797584	0.857755	-2.674783	0.068922	Yes
21		-1.52265	0.865041	-0.390311	0.676846	No Data
22		-3.280736	0.846293	-2.172938	0.113843	No Data
23		.	.	.	.	No Data
24	0.000416	-2.500991	0.87791	-1.351806	0.258772	Yes
25		-3.797584	0.857755	-2.674783	0.068922	No Data

Counts

8 Yes

2 No

80.00% Obs Within PB

Appendix H – Model B Validation Results

Estimating B5 Model

Validation	CG - Est (actual)	Predicted LN - Est	StdErr Indiv LN - Est	Est B5 80% UB	Back-T Est B5 80%	Actual within PB?
1	0.375279	-1.710195	0.783303	-0.688768	0.502194	Yes
2		-0.630784	0.781725	0.388586	1.474894	No Data
3		.	.	.	.	No Data
4	0.117899	-0.679015	0.802652	0.367643	1.444326	Yes
5	0.126393	-0.958136	0.768712	0.044265	1.045259	Yes
6	0.020228	-1.206511	0.779912	-0.189505	0.827368	Yes
7	0.002616	-2.621745	0.803697	-1.573724	0.207272	Yes
8	0.021295	-1.400352	0.797057	-0.36099	0.696986	Yes
9		.	.	.	.	No Data
10	0.127064	-1.449161	0.762222	-0.455223	0.634307	Yes
11	0.113038	-2.207531	0.784706	-1.184274	0.305968	Yes
12	0.128179	-2.229841	0.786328	-1.20447	0.299851	Yes
13	0.17266	.	.	.	.	No Data
14		.	.	.	.	No Data
15		-0.193133	0.809043	0.86186	2.367559	No Data
16	0.056279	-2.343332	0.789689	-1.313578	0.268856	Yes
17		-0.827515	0.794223	0.208152	1.2314	No Data
18	0.468366	.	.	.	.	No Data
19	0.103709	-2.10749	0.785515	-1.083178	0.338518	Yes
20		.	.	.	.	No Data
21		-0.773077	0.818824	0.294669	1.342682	No Data
22	0.040263	-0.814615	0.781138	0.203989	1.226285	Yes
23		-2.019325	0.836534	-0.928485	0.395152	No Data
24	0.394305	-0.535374	0.784926	0.48817	1.629332	Yes
25		-1.343795	0.765565	-0.345498	0.707868	No Data

Counts	
13	Yes
0	No
100.00%	Obs Within PB

Bibliography

Air Force Materiel Command. *AFMC Financial Management Handbook*. Wright-Patterson AFB OH: HQ AFMC, December 2001.

Calcutt, Harry M. *Cost Growth in DoD Major Programs: A Historical Perspective*. Executive Research Project. The Industrial College of the Armed Forces, National Defense University. Fort McNair Washington, D.C. 1993.

Christensen, David and Carl Templin. "EAC Evaluation Methods: Do They Still Work?," *Acquisition Review Quarterly*, 105 - 116 (Spring 2002).

Coleman, R. L., J. R. Summerville, M. DuBois, and B. Myers. "Risk in Cost Estimating: General Introduction & The BMDO Approach." Briefing at the 33rd Annual DoD Cost Analysis Symposium. Williamsburg VA. 2 February 2000.

Dameron, M. E., C. L. Pullen, J. R. Summerville, R. L. Coleman, and D. M. Snead. "NAVAIR Cost Growth: Overview of Analysis." Briefing at the Aeronautical Systems Center Industry Cost and Schedule Workshop. Wright-Patterson AFB OH. 24 April 2001.

Department of Defense. *Department of Defense Instruction Operation of the Defense Acquisition System*. DoD 5000.2. Washington: GPO, January 2001. Online Defense Acquisition Reference. http://web2.deskbook.osd.mil/reflib/MDOD/012DIDOC.HTM. 3 July 2002.

Drezner, J. A., J. M. Jarvaise, R. W. Hess, P. G. Hough, and D. Norton. *An Analysis of Weapon System Cost Growth*. Santa Monica CA: RAND, 1993 (MR-291-AF).

Garson, David. *PA 765 Statnotes: An Online Textbook.* Textbook for a statistics class taught at North Carlolina State University, Raleigh NC. http://www2.chass.ncsu.edu/garson/pa765/logistic.htm. 8 Jan 2003

Gordon, James A. Does A Rubber Baseline Guarantee Overruns? MS thesis, AFIT/GSM/LAS/96S-5. School of Systems and Logistics, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, September 1996 (AAA-7414).

Grossman, Elaine. "Defense Budgets Include $7.4 Billion Boost For Realistic Costing." Early Bird Article. Inside the Pentagon, February 7, 2002. https://www.asc.wpafb.af.mil/news/earlybird

Hough, Paul G. *Pitfalls in Calculating Cost Growth from Selected Acquisition Reports*. Santa Monica CA: RAND, 1992 (N-3136-AF).

Jarvaise, J. M., J. A. Drezner, D. Norton. *The Defense System Cost Performance Database: Cost Growth Analysis Using Selected Acquisition Reports*. Santa Monica CA: RAND, 1996 (MR-625-OSD).

Jaynes, Howard R, Jr. *Correlation Analysis: Army Acquisition Program Cycle Time and Cost Variation*. MS Thesis. Naval Postgraduate School. Monterey CA. September 1999.

JMP® Version 4.0.4, (Academic), CD-ROM. Computer software. SAS Institute Inc., Cary NC, 2001.

JMP® Version 5.0, (Academic), CD-ROM. Computer software. SAS Institute Inc., Cary NC, 2002

Knoche, Chris. "Defense Acquisition Deskbook: Selected Acquisition Report." Online Defense Acquisition Reference Guide. http://www.deskbook.osd.mil/default.asp. 1 July 2002.

Neter, John, Michael H. Kutner, Christopher J. Nachtsheim, and William Wasserman. *Applied Linear Statistical Models*. Boston: McGraw-Hill, 1996.

Rodgers, Chris and Jason Petrch. *Uncertainty & Risk Analysis.* A Practical Guide from Business Dynamics. PricewaterhouseCoopers Pamphlet. April 1999.

Sipple, Vincent P. *Estimating Engineering Cost Risk Using Logistic and Multiple Regression*. MS thesis, AFIT/GAQ/ENC/02-02. Graduate School of Engineering and Mangement, Air Force Institute of Technology (AU), Wright-Patterson AFB OH, March 2002 (ADA-400576).

Weinberger, Sharon. "Programs Must Use Realistic Cost Estimates, DoD Comptroller Says." Aerospace Daily, February 27, 2002. https://www.asc.wpafb.af.mil/news/aero-daily/2002/2002-02-05/ARTICLE198406.html

REPORT DOCUMENTATION PAGE			*Form Approved* *OMB No. 074-0188*
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of the collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to an penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**			
1. REPORT DATE *(DD-MM-YYYY)* 09-03-2003	**2. REPORT TYPE** **Master's Thesis**		**3. DATES COVERED** *(From – To)* Jun 2002 – Mar 2003
4. TITLE AND SUBTITLE ESTIMATING ENGINEERING AND MANUFACTURING DEVELOPMENT COST RISK USING LOGISTIC AND MULTIPLE REGRESSION			**5a. CONTRACT NUMBER**
			5b. GRANT NUMBER
			5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Bielecki, John, V., Captain, USAF			**5d. PROJECT NUMBER**
			5e. TASK NUMBER
			5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAMES(S) AND ADDRESS(S) Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 P Street, Building 640 WPAFB OH 45433-7765			**8. PERFORMING ORGANIZATION REPORT NUMBER** AFIT/GCA/ENC/03-01
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) ASC/FMCE, AFMC Attn: Mr. Michael J. Seibel 1865 4th St Rm 134 DSN: 986-5478 WPAFB OH 45433-7123 e-mail: Michael.Seibel@wpafb.af.mil			**10. SPONSOR/MONITOR'S ACRONYM(S)**
			11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.			
13. SUPPLEMENTARY NOTES			

14. ABSTRACT

This research effort validates, and further explores the use of a two-step procedure for assessing DoD major weapon system cost growth using historical data. We compile programmatic data from the Selected Acquisition Reports (SARs) between 1990 and 2001 for programs covering all defense departments. Our analysis concentrates on cost growth in the research and development dollar accounts for the Engineering and Manufacturing Development phase of acquisition. We investigate the use of logistic regression in cost growth analysis to predict whether or not cost growth will occur in a program. If applicable, a multiple regression step is implemented to predict how much cost growth will occur. Our study focuses on four of the seven SAR cost growth categories within the research and development accounts - schedule, estimating, support, and other. We investigate each of these four categories individually for significant cost growth characteristics and develop predictive models where appropriate.

15. SUBJECT TERMS

Logistic Regression, Multiple Regression, Cost Variance, Cost Growth, Selected Acquisition Report, SAR, DoD Cost Growth, Inferential Statistics, Cost Growth in DoD Acquisition Programs, Predicting Cost Growth

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON** Edward D. White, USAF, (ENC)
a. REPORT U	**b. ABSTRACT** U	**c. THIS PAGE** U	UU	98	**19b. TELEPHONE NUMBER** *(Include area code)* (937) 255-3636; e-mail: Edward.White@AFIT.edu

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39-18